

受限玻尔兹曼机的稀疏化特征学习

康丽萍^{1,2} 许光奎¹ 孙 显¹

(中国科学院电子学研究所 北京 100190)¹ (中国科学院大学 北京 100190)²

摘 要 受限玻尔兹曼机(RBM)作为深度学习算法的一种基础模型被广泛应用,但传统 RBM 算法没有充分考虑数据的稀疏化特征学习,使得算法性能受数据集的稀疏性影响较大。提出一种 RBM 稀疏化特征学习方法(sRBM),通过归一化的输入数据均值确定数据集的稀疏系数,将稀疏系数大于阈值的稠密数据集自动转化为稀疏数据集,在不损失信息量的情况下实现输入数据的稀疏化。在手写字符数据集和自然图像数据集上的实验结果表明,sRBM 通过输入数据稀疏化有效提升了 RBM 的稀疏化特征学习性能。

关键词 受限玻尔兹曼机(RBM),稀疏化,特征学习,置信网络,稳定性

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.12.016

Sparse Feature Learning for Restricted Boltzmann Machine

KANG Li-ping^{1,2} XU Guang-luan¹ SUN Xian¹

(Institute of Electronic, Chinese Academy of Sciences, Beijing 100190, China)¹

(University of Chinese Academy of Sciences, Beijing 100190, China)²

Abstract As a basic model for deep learning algorithms, restricted boltzmann machine (RBM) is widely applied in the field of machine learning. However, the traditional RBM algorithm does not take full account of the sparse feature learning for data. Therefore, the algorithm performance is greatly influenced by the sparsity of the dataset. In this study, a sparse feature learning method for restricted Boltzmann machine(sRBM) was proposed. Firstly, the sparse coefficient of the dataset is determined by the mean of normalized input data. Then the dense dataset with the sparse coefficient being greater than threshold will be converted to sparse dataset automatically. As a result, sRBM makes the input data sparse without information loss. We performed experiments on MNIST dataset and attribute discovery dataset. The experimental results show that sRBM improves the performance of sparse feature learning for RBM effectively.

Keywords Restricted Boltzmann machine(RBM), Sparsity, Feature learning, Belief network, Stability

1 引言

随着互联网技术的发展,数据量呈爆炸式增长,机器学习利用计算机从大量数据中挖掘规律、模仿人类行为,是实现数据信息自动学习的有力技术手段,为解决众多领域的不同任务提供了途径。数据的特征学习作为机器学习系统的输入单元,对模型的构建至关重要,模型性能的优劣在很大程度上取决于学习到的数据特征质量。好的特征应针对特定的任务保留有用的数据信息,尽可能去除数据冗余;此外,特征的稀疏化表示能够降低数据间的相关性,从而抑制模型的过拟合现象,提高系统的鲁棒性。稀疏化的特征表示无论是在理论研究还是实际应用中都具有很大优势^[1-3],因此稀疏化特征学习是提升机器学习模型性能的重要手段。

非监督模型能够自动学习数据特征,在数据特征学习方面发挥了重要作用,广泛应用于目标检测、识别、预测和可视化任务中。同时,通过堆叠无监督模型产生多层特征表达,并结合传统监督分类方法得到的深度学习结构已经成为机器学

习领域的一个研究热点^[4,5]。基于能量模型的受限玻尔兹曼机(Restricted Boltzmann Machine, RBM)^[6]作为当前深度学习算法的一种基础模型,以其神经网络结构和快速学习算法在数据降维^[7]、语音识别^[8]、3D 物体识别^[9]、图像转换^[10]以及高维时间序列建模^[11]等机器学习领域均获得了广泛应用。然而,传统的 RBM 算法没有充分考虑数据的稀疏化特征学习,使得 RBM 算法的性能受数据集的稀疏性影响较大,限制了 RBM 算法的应用。

目前研究者们主要从 RBM 隐层单元的稀疏惩罚方面开展有关 RBM 稀疏性问题的研究。Lee 等人^[12]提出了基于误差平方和的稀疏惩罚因子,实现了 RBM 隐层单元偏置的稀疏化;Hinton 等人^[6]提出了基于交叉熵的稀疏惩罚因子,实现了 RBM 权值和隐层单元偏置的稀疏化;Luo 等人^[13]通过引入隐单元激活概率的混合范数,构建了稀疏组受限玻尔兹曼机(SGRBM),降低了特征之间的相关性;Guo 等人^[14]通过增加范数调节因子对 SGRBM 算法进行改进。隐层单元稀疏惩罚方法虽然能够实现 RBM 内部结构单元的稀疏化,但并

到稿日期:2016-01-07 返修日期:2016-04-15 本文受国家自然科学基金(41301493),高分对地观测领域学术交流项目(GFEX04060103)资助。
康丽萍(1991—),女,硕士生,主要研究方向为图像处理、信息融合与数据挖掘,E-mail: kangliping13@mails.ucas.ac.cn;许光奎(1978—),男,博士,研究员,主要研究方向为图像处理、计算机视觉与数据挖掘;孙 显(1981—),男,博士,副研究员,主要研究方向为图像分析、理解与信息融合。

没有考虑到数据集稀疏性的影响,在稀疏化程度不同的数据集上的算法性能可能存在较大差异。

针对现阶段研究存在的问题,本文提出一种 RBM 稀疏化特征学习方法(sRBM)。该方法通过归一化的训练数据均值定义稀疏系数,对于稀疏系数高于阈值的稠密数据集,将其自动转化为对应的稀疏数据集,在不损失信息量的前提下实现输入数据的稀疏化,提高了 RBM 算法的稀疏化特征学习性能。

2 RBM 稀疏化特征学习方法

RBM^[6]是一种无监督的学习模型,具有两层网络结构:一层为可视层,另一层为隐藏层,两层之间的联合概率满足玻尔兹曼分布。使用能量函数作为整个网络状态的测度,当网络最稳定时能量函数值最小,RBM 的能量函数为:

$$E(v, h; \theta) = -\sum_{i=1}^V \sum_{j=1}^H v_i h_j w_{ij} - \sum_{i=1}^V v_i b_i^v - \sum_{j=1}^H h_j b_j^h \quad (1)$$

其中, V 和 H 分别为可视层和隐藏层的节点数目, v_i 和 h_j 分别表示可视层第 i 个节点和隐藏层第 j 个节点的状态, $\theta = \{w_{ij}, b_i^v, b_j^h\}$ 是 RBM 算法的参数, w_{ij} 为连接可视层第 i 个节点和隐藏层第 j 个节点的权重值, b_i^v 为可视层第 i 个节点的偏置项, b_j^h 为隐藏层第 j 个节点的偏置项。

RBM 作为一种特征自动学习模型,特征学习的质量直接影响其模型性能,尤其是稀疏化的特征表示能够降低数据冗余,提高系统鲁棒性。将图像训练数据集的稀疏系数定义为 s :

$$s = \frac{1}{N \cdot K} \sum_{n=1}^N \sum_{i=1}^K v_i^n \quad (2)$$

其中, v_i^n 表示第 n 个训练样本第 i 维的数值, $v_i^n \in [0, 1]$, N 为训练样本的数目, K 为样本维度。当 $s \geq \alpha$ 时将训练数据集定义为稠密数据集,当 $s < \alpha$ 时定义为稀疏数据集, α 为稀疏性阈值,对于归一化到 0~1 的数据通常选取 $\alpha = 0.5$ 。

对稀疏系数为 s 的图像训练数据集,将其数据值反转可以得到反转图像训练数据集(Negative Training Images),对应的稀疏系数为 s_N 。由于 $s + s_N = 1$,因此在不损失图像信息的情况下,通过数据值反转可以将稠密数据集自动转化为稀疏数据集。

RBM 算法对稀疏化程度不同的数据集的特征学习能力不同,为了降低数据集稀疏性对 RBM 特征学习性能的影响,需要对输入数据 v_i 进行稀疏化。当 $s \geq \alpha$ 时,输入为稠密数据集,可通过数据值反转实现输入数据集稀疏化;当 $s < \alpha$ 时,输入为稀疏数据集,保持不变。定义输入数据 v_i 的稀疏化表达 z_i :

$$z_i = \left(\frac{1}{2} - v_i \right) \text{sign}(s - \alpha) + \frac{1}{2} = \begin{cases} 1 - v_i, & s \geq \alpha \\ v_i, & s < \alpha \end{cases} \quad (3)$$

由式(1)和式(3)可知,sRBM 算法的能量函数为:

$$E(z, h; \theta) = -\sum_{i=1}^V \sum_{j=1}^H z_i h_j w_{ij} - \sum_{i=1}^V z_i b_i^v - \sum_{j=1}^H h_j b_j^h \quad (4)$$

采用 sigmoid 函数作为激活函数,在给定可视层节点状态时,第 j 个隐藏层节点激活的概率为:

$$P(h_j = 1 | z) = \frac{1}{1 + \exp(-\sum_i z_i w_{ij} - b_j^h)} \quad (5)$$

同理,第 i 个可视层节点激活的概率为:

$$P(z_i = 1 | h) = \frac{1}{1 + \exp(-\sum_j h_j w_{ij} - b_i^v)} \quad (6)$$

令 $C = \frac{1}{2} (1 + \text{sign}(s - \alpha))$, 则

$$\begin{aligned} z_i &= \left(\frac{1}{2} - v_i \right) \text{sign}(s - \alpha) + \frac{1}{2} \\ &= -\text{sign}(s - \alpha) v_i + \frac{1}{2} (1 + \text{sign}(s - \alpha)) \\ &= -\text{sign}(s - \alpha) v_i + C \end{aligned} \quad (7)$$

将式(7)代入式(4)可将 sRBM 算法的能量函数化简为:

$$\begin{aligned} E(z, h; \theta) &= -\sum_{i=1}^V \sum_{j=1}^H z_i h_j w_{ij} - \sum_{i=1}^V z_i b_i^v - \sum_{j=1}^H h_j b_j^h \\ &= -\sum_{i=1}^V \sum_{j=1}^H (-\text{sign}(s - \alpha) v_i + C) h_j w_{ij} - \sum_{i=1}^V (-\text{sign}(s - \alpha) v_i + C) b_i^v \\ &= -\sum_{i=1}^V \sum_{j=1}^H (-\text{sign}(s - \alpha) v_i) h_j w_{ij} - \sum_{i=1}^V (-\text{sign}(s - \alpha) v_i) b_i^v - \sum_{j=1}^H (b_j^h + \sum_{i=1}^V C w_{ij}) h_j + \text{const} \end{aligned} \quad (8)$$

利用最大似然法估计 sRBM 的参数,其对数似然函数为:

$$L(\theta) = \sum_{t=1}^T (\log \sum_{h,v} \exp(-E(z^{(t)}, h; \theta)) - \log \sum_{v,h} \exp(-E(z, h; \theta))) \quad (9)$$

其中, T 为训练数据集的样本数目。

利用 CD 算法^[15]计算得到 sRBM 的参数更新规则如下:

$$w_{ij} = w_{ij} + (-\text{sign}(s - \alpha)) \eta_w (\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}) \quad (10)$$

$$b_i^v = b_i^v + (-\text{sign}(s - \alpha)) \eta_v (\langle v_i \rangle_{data} - \langle v_i \rangle_{model}) \quad (11)$$

$$b_j^h = b_j^h + \eta^h (\langle h_j \rangle_{data} - \langle h_j \rangle_{model}) \quad (12)$$

其中, η 为学习率, $\langle \cdot \rangle_p$ 表示关于分布 p 的均值。

由式(8)可知, b_j^h 在 sRBM 算法迭代完成后需要更新为:

$$b_j^h = b_j^h + \sum_{i=1}^V C w_{ij} \quad (13)$$

sRBM 算法训练过程的具体步骤如下。

算法 sRBM 算法

输入: 图像像素值向量 v_i

算法步骤:

0. 利用正态分布随机初始化 w_{ij} , 将 b_i^v 和 b_j^h 初始化为 0

for $t = 1$: NumberEpochs do

for $n = 1$: NumberBatches do

正向阶段:

1. 计算输入向量 v_i 的稀疏化表达

$$z_i = \left(\frac{1}{2} - v_i \right) \text{sign}(s - \alpha) + \frac{1}{2}$$

2. 利用式(5)计算隐层节点的激活状态

3. 计算正向传播阶段的最大似然估计(MLE)梯度

负向阶段:

4. 利用式(6)计算 z_i 的重构值 $\hat{z}_i$

5. 计算 v_i 的重构值 $\hat{v}_i = \frac{1}{2} - \text{sign}(s - \alpha) (\hat{z}_i - \frac{1}{2})$

6. 利用式(5)计算隐层节点的激活状态

7. 计算反向传播阶段的 MLE 梯度

参数更新:

8. 利用式(10)一式(12)分别对 w_{ij} , b_i^v 和 b_j^h 进行更新

end for

end for

9. 更新隐层偏置项 $b_j^h \leftarrow b_j^h + \sum_{i=1}^V C w_{ij}$

输出:

sRBM 算法参数 $\theta = \{w_{ij}, b_i^v, b_j^h\}$

3 基于 sRBM 的置信网络

本文结合 softmax 多分类器和 BP 算法构建基于 sRBM 的置信网络,作为图像分类的基本框架。本文利用该置信网络从分类精确度的角度对 sRBM 算法稀疏化特征学习性能进行定量实验分析。

置信网络首先利用 sRBM 算法进行图像的稀疏化特征学习,再利用 softmax 多分类器^[16]对学习到的特征进行多分类,softmax 计算公式为:

$$p_i = p(y=i|x;\theta) = \frac{e^{\theta_i^T x}}{\sum_{j=1}^k e^{\theta_j^T x}} \quad (14)$$

其中, x 和 y 分别代表图像实例及其对应的类别, p_i 代表该图像实例判定为类别 i 的概率, $i \in \{1, 2, \dots, k\}$, k 为数据集包含的类别数目, θ 为优化参数。

完成 sRBM 特征学习和多分类任务后,使用 BP 算法^[17]对网络模型参数、输出层偏置、隐层偏置和权值矩阵进行全局优化,为使模型的实际输出尽可能接近期望输出,误差函数使用交叉熵:

$$H(r, p) = - \sum_{i=1}^k (r_i \log p_i + (1-r_i) \log(1-p_i)) \quad (15)$$

其中, $r \in \{0, 1\}^k$ 代表数据的真实分类,假设真实类别为 i 时, $r_i = 1, r_{-i} = 0$ 。利用最小化误差函数对网络参数 θ 进行训练:

$$\theta^* = \arg \min_{\theta} H(r, p) \quad (16)$$

4 实验结果与分析

4.1 实验数据及评价标准

采用 MNIST 手写字符数据集和 Attribute Discovery 自然图像数据集作为实验数据集对本文提出的 sRBM 稀疏化特征学习性能进行实验分析,并与传统 RBM^[6]进行分析比较。

MNIST 数据集^[18]为手写字符数据集,包含 10 个数字类别,共 70000 幅图像,随机选取 60000 幅作为训练集,10000 幅作为测试集,并重复 5 次得到 5 组实验数据。训练集平均稀疏系数 $s=0.1309$,为稀疏数据集;通过对 MNIST 数据集进行数据值反转可以得到平均稀疏系数 $s_N=0.8691$ 的稠密数据集。分别在 MNIST 数据集对应的稀疏和稠密数据集上对 sRBM 和 RBM 算法的稀疏化特征学习性能进行对比实验。

Attribute Discovery 数据集^[19]为自然图像数据集,包含 27 个购物类别,共 37794 幅图像,随机选取 30000 幅作为训练集,7700 幅作为测试集,并重复 5 次得到 5 组实验数据。训练集平均稀疏系数 $s=0.8296$,为稠密数据集;通过对 Attribute Discovery 数据集进行数据值反转可以得到平均稀疏系数 $s_N=0.1704$ 的稀疏数据集,分别在 Attribute Discovery 数据集对应的稀疏和稠密数据集上对 sRBM 和 RBM 算法的稀疏化特征学习性能进行对比实验。

sRBM 的图像稀疏化特征学习性能一方面可以通过可视化权重值 w 对 sRBM 学习到的特征滤波器质量进行定性分析;另一方面由于 sRBM 学习到的图像特征质量直接影响基

于 sRBM 的置信网络分类精确度,因此通过比较其对应置信网络的分类精确度对 sRBM 的稀疏化特征学习性能进行定量分析。假设测试数据共 M 幅,其中分类正确的数目为 M_T 幅,则分类精确度 p 定义为:

$$p = \frac{M_T}{M} \times 100\% \quad (17)$$

4.2 实验结果

将传统 RBM 算法^[6]作为基准方法与本文提出的 sRBM 算法进行分析比较,分别在 MNIST 和 Attribute Discovery 数据集对应的 5 组稀疏和稠密数据集上对 sRBM 和 RBM 算法图像稀疏化特征学习性能进行对比实验。

一方面对比了 sRBM 和 RBM 算法学习到的特征滤波器质量,并随机选取一组实验结果进行可视化分析;另一方面为了定量分析 sRBM 和 RBM 算法的特征学习性能,分别对基于 sRBM 和 RBM 算法构建的置信网络在 5 组实验数据上分类迭代 2 到 30 次时的平均分类精确度进行实验对比。

实验中 sRBM 和 RBM 算法的参数依据文献^[6]中的通用准则进行选取,在其合理范围内随机选定一组参数进行实验,并对 sRBM 和 RBM 算法对参数的敏感度进行实验分析。

4.2.1 MNIST 数据集

MNIST 数据集对应的稀疏和稠密数据集如图 1 所示。

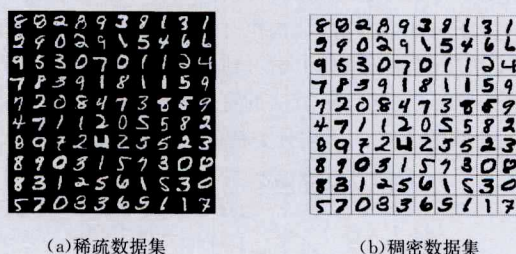


图 1 MNIST 数据集

依据参数选取通用规则,sRBM 和 RBM 均采用 500 个隐节点,0.01 的学习率,30 次迭代,利用 softmax 多分类方法进行图像分类,利用 BP 算法进行参数优化,并分类迭代 30 次。

1) MNIST 稀疏数据集

利用 RBM 和 sRBM 算法在 MNIST 稀疏数据集上学习得到的特征滤波器如图 2 所示。

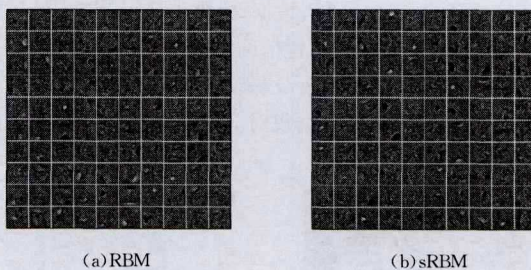


图 2 MNIST 稀疏数据集特征滤波器对比

基于 RBM 和 sRBM 算法的置信网络在 5 组数据上分类迭代 2 到 30 次对应的平均分类精确度变化如图 3 所示。

由图 2 和图 3 对比发现,本文提出的 sRBM 算法与 RBM 算法在 MNIST 稀疏数据集上学习得到的特征滤波器质量和对应的置信网络分类精确度基本相同。这是由于 MNIST 稀疏数据集的稀疏性系数 $s=0.1309$,sRBM 算法保持原始图像数据值不变,因此 sRBM 和 RBM 算法在 MNIST 稀疏数据集上的图像稀疏化特征学习性能相当。

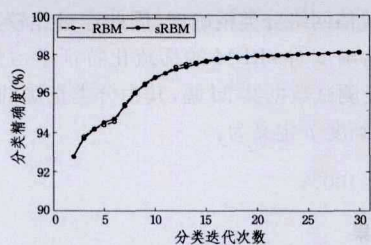


图3 MNIST 稀疏数据集分类精度对比

2) MNIST 稠密数据集

利用 RBM 和 sRBM 算法在 MNIST 稠密数据集上学习得到的特征滤波器如图 4 所示。

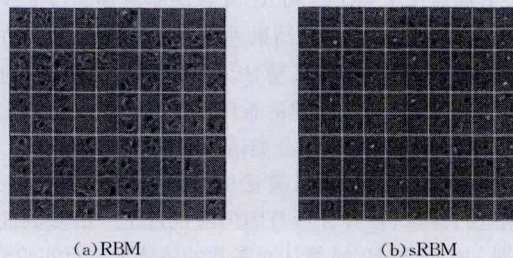


图4 MNIST 稠密数据集特征滤波器对比

由图 4 对比可知, sRBM 算法在 MNIST 稠密数据集上学习到的特征信息更丰富, 结构更显著, 明显优于 RBM 算法, 且与 MNIST 稀疏数据集上学习到的图像特征质量相当。

基于 RBM 和 sRBM 算法的置信网络在 5 组数据上分类迭代 2 到 30 次对应的平均分类精度变化如图 5 所示。

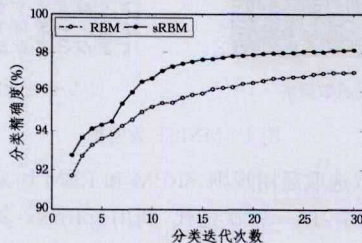


图5 MNIST 稠密数据集分类精度对比

由图 5 对比发现, 在 MNIST 稠密数据集上基于 sRBM 算法得到的分类精度明显高于 RBM 算法, 进一步验证了 sRBM 算法通过稀疏化输入数据, 改善了 RBM 算法的稀疏化特征学习能力, 进而提高了置信网络的分类精度。

4.2.2 Attribute Discovery 数据集

Attribute Discover 数据集对应的稀疏和稠密数据集如图 6 所示。

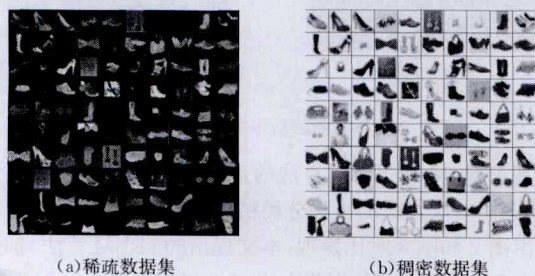


图6 Attribute Discovery 数据集

依据参数选取通用规则, sRBM 和 RBM 均采用 700 个隐节点, 0.1 的学习率, 20 次迭代, 利用 softmax 多分类方法进行图像分类, 利用 BP 算法进行参数优化, 分类迭代 30 次。

1) Attribute Discovery 稀疏数据集

利用 RBM 和 sRBM 算法在 Attribute Discovery 稀疏数据集上学习得到的特征滤波器如图 7 所示。

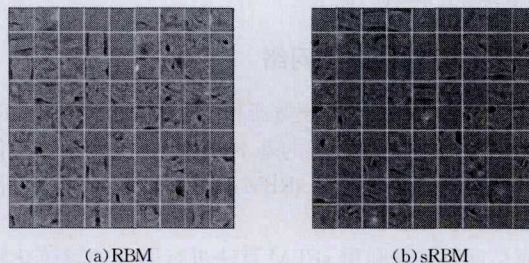


图7 Attribute Discovery 稀疏数据集特征滤波器对比

基于 RBM 和 sRBM 算法的置信网络在 5 组数据上分类迭代 2 到 30 次对应的平均分类精度变化如图 8 所示。

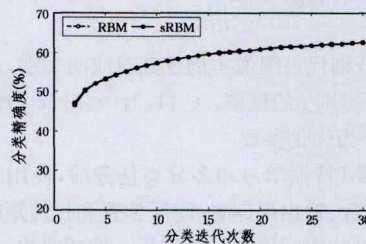


图8 Attribute Discovery 稀疏数据集分类精度对比

Attribute Discovery 稀疏数据集的稀疏系数 $s=0.1704$, 由图 7 和图 8 对比发现本文提出的 sRBM 算法与 RBM 算法在 Attribute Discovery 稀疏数据集上的图像稀疏化特征学习性能相当。

2) Attribute Discovery 稠密数据集

利用 RBM 和 sRBM 算法在 Attribute Discovery 稠密数据集上学习得到的特征滤波器如图 9 所示。

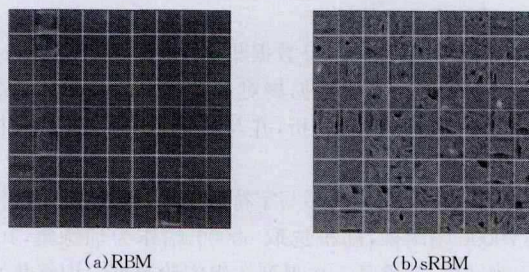


图9 Attribute Discovery 稠密数据集特征滤波器对比

将图 9 与图 7 进行对比可知, RBM 算法在 Attribute Discovery 稠密数据集上学习到的图像特征质量相比 Attribute Discovery 稀疏数据集明显变差, 而 sRBM 算法则基本不变。sRBM 算法的稀疏化特征学习性能在 Attribute Discovery 稠密数据集上明显优于 RBM 算法。

基于 RBM 和 sRBM 算法的置信网络在 5 组数据上分类迭代 2 到 30 次对应的平均分类精度变化如图 10 所示。

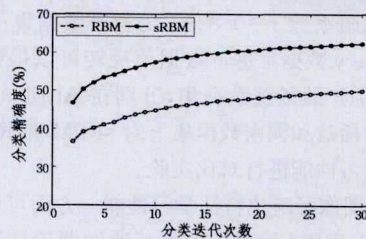


图10 Attribute Discovery 稠密数据集分类精度对比

由图 10 对比可以发现,基于 sRBM 算法的置信网络在 Attribute Discovery 稠密数据集上得到的分类精确度相比 RBM 算法有了较大提升,进一步验证了 sRBM 算法在稠密数据集中通过稀疏化输入数据增强了图像稀疏化特征学习性能,提高了置信网络的分类精确度。

为了进一步比较 RBM 和 sRBM 算法的鲁棒性,分别计算了 5 组实验数据对应分类精确度的标准方差,结果如表 1 所列(σ_s 对应稀疏数据集, σ_d 对应稠密数据集)。

表 1 RBM 和 sRBM 算法鲁棒性比较

算法对比	MNIST 数据集		Attribute Discovery 数据集	
	σ_s	σ_d	σ_s	σ_d
RBM	0.0896	0.2988	0.7943	1.1079
sRBM	0.0736	0.1167	0.6985	0.6063

由表 1 可知,RBM 和 sRBM 算法在 MNIST 和 Attribute Discovery 数据集对应的稀疏和稠密数据集上的分类性能均比较稳定,且 sRBM 算法相比 RBM 算法鲁棒性更强。

参数敏感度对算法性能有重要影响,本文主要针对隐层节点数目和学习率两个参数进行实验并分析了 RBM 和 sRBM 对输入参数的敏感度,结果如表 2 所列。

表 2 RBM 和 sRBM 算法对输入参数的敏感度比较

数据集	序号	隐层节点数目	学习率	RBM		sRBM	
				P_s	P_d	P_s	P_d
MNIST	1	500	0.01	0.9804	0.9695	0.9807	0.9797
	2	500	0.03	0.9821	0.9726	0.9823	0.9818
	3	500	0.05	0.9834	0.9704	0.9836	0.9849
	4	300	0.01	0.9793	0.9616	0.9801	0.9772
	5	500	0.01	0.9804	0.9695	0.9807	0.9797
	6	700	0.01	0.9813	0.9733	0.9806	0.9813
Attribute Discovery	1	700	0.01	0.5847	0.5456	0.5865	0.5866
	2	700	0.05	0.6096	0.5625	0.6137	0.6084
	3	700	0.10	0.6213	0.4870	0.6252	0.6248
	4	300	0.10	0.6084	0.5008	0.6104	0.6040
	5	500	0.10	0.6232	0.4842	0.6197	0.6225
	6	700	0.10	0.6213	0.4870	0.6252	0.6248

对比表 2 的实验结果可知,在隐层节点数目固定、学习率变化的情况下,RBM 和 sRBM 算法在 MNIST(或 Attribute Discovery)稀疏数据集上的分类精确度相当,在 MNIST(或 Attribute Discovery)稠密数据集上 sRBM 算法的分类精确度均高于 RBM 算法,表明 sRBM 相比 RBM 通过对输入数据稀疏化带来的分类性能提升不受学习率变化的影响。同理可知,sRBM 相比 RBM 通过对输入数据稀疏化带来的分类性能提升也不受隐层节点数目变化的影响。

由实验结果可知,sRBM 相比 RBM 在特征学习性能方面对数据集稀疏性的依赖程度更低,本文将稳定性作为量化指标,对比分析了 sRBM 和 RBM 算法对输入数据稀疏性的依赖程度,稳定性 St 定义为:

$$St = \frac{1}{1 + \gamma \cdot |p_s - p_d|} \quad (18)$$

其中, p_s 和 p_d 分别指置信网络在同一数据集对应的稀疏和稠密数据集上的平均分类精确度, γ 为稳定性系数,本文选取 $\gamma=100$ 。

St 越大,算法稳定性越高,对输入数据稀疏性的依赖程度越低;反之,算法稳定性越低,对输入数据稀疏性的依赖程度越高。RBM 和 sRBM 算法在 Attribute Discovery 和 MNIST 数据集上对于输入数据稀疏性的稳定性如表 3 所列。

表 3 RBM 和 sRBM 算法对输入数据稀疏性的稳定性比较

算法对比	MNIST 数据集			Attribute Discovery 数据集		
	P_s	P_d	St	P_s	P_d	St
RBM	0.9809	0.9699	0.48	0.6256	0.4969	0.07
sRBM	0.9815	0.9813	0.98	0.6246	0.6237	0.92

由表 3 对比可以发现,sRBM 算法在 MNIST 数据集和 Attribute Discovery 数据集上对于输入数据稀疏性的稳定性均高于 RBM 算法。由此可见,sRBM 算法在稀疏数据集和稠密数据集上均学习到较好的数据特征,改善了 RBM 算法受数据集稀疏性的影响,有效提升了其稀疏化特征学习性能。

结束语 本文提出一种 RBM 的稀疏化特征学习方法(sRBM),与传统 RBM 算法相比,从输入数据稀疏化的角度增强了 RBM 算法的稀疏化特征学习性能。首先依据归一化的训练数据均值确定该数据集的稀疏系数,将稀疏系数大于阈值的稠密数据集转化为稀疏数据集,在不损失信息量的前提下实现了输入数据的稀疏化。基于 MNIST 和 Attribute Discovery 数据集对应的稀疏和稠密数据集,从可视化特征学习滤波器和对应置信网络分类精确度两个方面对 sRBM 和 RBM 算法进行了对比实验。实验发现,sRBM 算法无论在字符数据集还是自然图像数据集对应的稠密数据集中,其学习到的特征滤波器质量及对应置信网络的分类精确度均高于 RBM 算法,且对输入数据的稀疏性具有更高的稳定性。结果表明,sRBM 算法通过稀疏化输入数据有效提升了 RBM 算法在稠密数据集中的稀疏化特征学习性能。今后工作将尝试从输入数据稀疏化和隐层单元稀疏化两个维度同时对 RBM 算法进行改进,进一步提升 RBM 算法的特征学习性能,并将其应用于更广泛的实际任务中。

参 考 文 献

[1] Poultney C, Chopra S, Cun Y L. Efficient Learning of Sparse Representations with an Energy-Based Model[C]//Proceedings of Advances in Neural Information Processing Systems, Vancouver, B. C; NIPS, 2006; 1137-1144

[2] Doi E, Balcan D C, Lewicki M S. A Theoretical Analysis of Robust Coding over Noisy Overcomplete Channels[C]//Proceedings of Advances in Neural Information Processing Systems, Vancouver, B. C; NIPS, 2006; 37-3311

[3] Olshausen B A, Field D J. Sparse Coding with an Overcomplete Basis Set; A Strategy Employed by V1? [J]. Vision Research, 1997, 37(23): 3311-3325

[4] Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(5786): 504-507

[5] Guo L L, Ding S F. 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

[6] Hinton G. A Practical Guide to Training Restricted Boltzmann Machines[J]. Momentum, 2010, 9(1): 599-619

[7] Chen Y, Zheng D Q, Zhao T J. Chinese relation extraction based on Deep Belief Nets[J]. Journal of Software, 2012, 23(10): 2572-2585(in Chinese)

陈宇, 郑德权, 赵铁军. 基于 Deep Belief Nets 的中文名实体关系抽取[J]. 软件学报, 2012, 23(10): 2572-2585

[8] Mohamed A R, Dahl G E, Hinton G. Acoustic Modeling Using Deep Belief Networks[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 14-22

[9] Schmäh T, Hinton G E, Small S L, et al. Generative Versus Dis-

criminative Training of Rbms for Classification of Fmri Images [C]// Proceedings of Advances in Neural Information Processing Systems. Vancouver, B. C. ; NIPS, 2008; 1409-1416

- [10] Memisevic R, Hinton G. Unsupervised Learning of Image Transformations[C]// Proceedings of Computer Vision and Pattern Recognition(CVPR'07). Minneapolis, MN; IEEE, 2007; 1-8
- [11] Taylor G W, Sigal L, Fleet D J, et al. Dynamical Binary Latent Variable Models for 3d Human Pose Tracking[C]// 2010 IEEE Conference on Proceedings of Computer Vision and Pattern Recognition(CVPR). San Francisco, CA; IEEE, 2010; 631-638
- [12] Lee H, Ekanadham C, Ng A Y. Sparse Deep Belief Net Model for Visual Area V2[C]// Proceedings of Advances in Neural Information Processing Systems. Vancouver, B. C; NIPS, 2008; 873-880
- [13] Luo H, Shen R, Niu C, et al. Sparse Group Restricted Boltzmann Machines[C]// Proceedings of AAAI Conference on Artificial Intelligence. Georgia, USA; AAAI, 2010
- [14] Guo R, Qi H. Partially-Sparse Restricted Boltzmann Machine for Background Modeling and Subtraction[C]// 2013 12th Interna-

tional Conference on Proceedings of Machine Learning and Applications (ICMLA). Miami, FL; IEEE, 2013; 209-214

- [15] Carreira-Perpinan M A, Hinton G E. On Contrastive Divergence Learning[C]// Proceedings of Proceedings of the tenth international workshop on Artificial Intelligence and Statistics. Barbados; AISTATS, 2005; 33-40
- [16] Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann Machines for Collaborative Filtering[C]// Proceedings of Proceedings of the 24th International Conference on Machine Learning. New York; ACM, 2007; 791-798
- [17] Pineda F J. Generalization of Back-Propagation to Recurrent Neural Networks[J]. Physical Review Letters, 1987, 59(19): 307-308
- [18] LeCun Y, Cortes C, Burges C J. The Mnist Database of Handwritten Digits[OL]. <http://yann.lecun.com/exdb/mnist>
- [19] Berg T L, Berg A C, Shih J. Automatic Attribute Discovery and Characterization from Noisy Web Data [C] // Proceedings of Computer Vision (ECCV 2010). Crete, Greece; ECCV, 2010; 663-676

(上接第 90 页)

同理可以得到元素 x_i 在其他时期的优势度表, 由式(2)得到属性权重 $O = \{0.2034, 0.3051, 0.2712, 0.2203\}$, 由式(3)得到各元素的平均优势得分表, 如表 8 所列。

表 8 平均优势得分表

	1	2	3	4
x_1	0.8488	0.8239	0.8822	0.8609
x_2	0.9432	0.9432	0.8978	0.9006
x_3	0.9204	0.9204	0.9381	0.9315
x_4	0.8533	0.8533	0.8888	0.8990
x_5	0.8595	0.8368	0.7914	0.8459

因此, 可以根据式(6)得到元素 x_i 的综合得分值: $C(x_1) = 0.8578, C(x_2) = 0.9124, C(x_3) = 0.9297, C(x_4) = 0.8823, C(x_5) = 0.8321$ 。

综上所述得出的排序如下:

$$C(x_3) > C(x_2) > C(x_4) > C(x_1) > C(x_5)$$

因此, 对于该公司而言, x_3 为担任公司经理的最佳人选, x_2 次之。

将本文得到的决策结果与文献[3]和文献[7]的决策结果进行对比, 对比结果如表 9 所列。

表 9 决策结果对比

	文献[3]	文献[7]	本文
决策结果	x_2	x_3	x_3
时间复杂度	$O(n^2)$	$O(n^2)$	$O(mn^2)$

由结果分析可知, 文献[3]和文献[7]的时间复杂度均为 $O(n^2)$, 但二者均没有考虑到属性权重对决策结果的影响, 默认各个属性的权重彼此相等, 这就使得所提出的算法与实际情况不符, 且二者均使用由水平软集得出的机会值作为最终决策结果的判定, 使产生的决策结果具有概率性。

本文的时间复杂度是 $O(mn^2)$, 充分利用已知的数据进行属性权重的计算, 弥补了依靠主观经验获得属性权重的缺陷, 使得决策方法更加适用于实际, 决策结果更加精确。

结束语 在经典模糊软集的基础上, 首先研究了在某一属性下的元素优势度问题, 得到某个时刻下的优势度表, 然后将该时刻下的优势度表和对应的属性权重相结合, 得到该时刻下元素的平均优势得分值, 最后将各个优势得分值和时间

权重结合, 得到综合得分值, 再将综合得分值进行比较, 即得到最终决策结果。通过实例对本文算法的正确性及可行性进行验证, 再与其他决策方法进行比较分析, 结果验证了该决策方法的合理性。

参 考 文 献

- [1] Molodtsov D. Soft set theory-first results [J]. Computers and Mathematics with Applications, 1999, 37(4/5): 19-31
- [2] Zadeh L A. Fuzzy Sets[J]. Information and Control, 1965, 8(3): 338-353
- [3] Mao J J, Yao D B, Wang C C. Group decision-making method based on time-series fuzzy soft sets [J]. Systems Engineering-Theory & Practice, 2014, 34(1): 181-189 (in Chinese)
- 毛建军, 姚登宝, 王翠翠. 基于时序模糊软集的群决策方法[J]. 系统工程理论与实践, 2014, 34(1): 181-189
- [4] Maji P K, Biswas R, Roy A R. Soft Sets Theory [J]. Computers and Mathematics with Applications, 2003, 45(4/5): 555-562
- [5] Ali M I, Feng F. On Some New Operations in Soft Set Theory [J]. Computers and Mathematics with Applications, 2009, (57): 1547-1553
- [6] Sezgin A, Atagun A S. On Operations of Soft Sets [J]. Computers and Mathematics with Applications, 2011, 61(5): 1457-1467
- [7] Qian Q Q, Wu T, et al. Group decision making method based on dynamic bipolar-value fuzzy soft sets [J]. Computer Engineering and Applications, 2014, 50(12): 38-41 (in Chinese)
- 钱庆庆, 吴涛, 等. 基于动态双极值模糊软集的群决策方法[J]. 计算机工程与应用, 2014, 50(12): 38-42
- [8] Pawlak Z. Rough sets [J]. International Journal of Computer & Information Sciences, 1982, 11(5): 341-356
- [9] Zadeh L A. Fuzzy sets [J]. Information and Control, 1965, 8(3): 338-353
- [10] Atanassov K. Intuitionistic fuzzy sets [J]. Fuzzy Sets and Systems, 1986, 20(1): 87-96
- [11] Liao B Q, Hu T Z. Probability theory tutorial (Version 2) [M]. University of Science and Technology of China Press, 2009 (in Chinese)
- 缪柏其, 胡太忠. 概率论教程(2版) [M]. 中国科学技术大学出版社, 2009