

doi:10.11918/j.issn.0367-6234.2016.05.025

改进卷积玻尔兹曼机的图像特征深度提取

刘 凯¹,张立民²,范晓磊³

(1.海军航空工程学院 基础实验部,264001 山东 烟台; 2. 海军航空工程学院 信息融合研究所, 264001 山东 烟台;
3. 第二炮兵工程大学 士官职业技术教育学院,262500 山东 青州)

摘 要: 针对卷积深度和信念网络存在计算复杂度高和训练缓慢的问题,提出卷积深度玻尔兹曼机用于图像特征提取. 针对卷积受限玻尔兹曼机进行改进,提出最大化图像中间区域概率的训练目标函数,并引入性能较好的交叉熵稀疏惩罚因子和 dropout 训练方法. 设计卷积深度玻尔兹曼机结构,提出均值聚合机制,将聚合层内点的值定义为 block 中各点激活概率均值,对层间关联进行简化,将聚合层内各面直接叠加以供高层 CRBM 提取特征. 通过在 MNIST 手写数字识别集上的实验结果证明,采用新模型提取的图像特征分类准确率提高 0.5%、训练时间减少 50%,且达到了目前 MNIST 数据集的最佳水平.

关键词: 深度学习;图像特征提取;卷积受限玻尔兹曼机;卷积深度玻尔兹曼机

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 0367-6234(2016)05-0155-05

New image deep feature extraction based on improved CRBM

LIU Kai¹, ZHANG Limin², FAN Xiaolei³

(1. Department of Basic Experiment, Naval Aeronautical and Astronautical University, 264001 Yantai, Shandong, China;
2. Research Institute of Information Fusion, Naval Aeronautical and Astronautical University, 264001 Yantai, Shandong, China;
3. Noncommissioned Officers Vocational and Technical Education College, The Second Artillery Engineering University, 262500 Qingzhou, Shandong, China)

Abstract: To resolve the problems of high computational complexity and slow training in Convolutional Deep Belief Net, Convolutional Deep Boltzmann Machine (CDBM) is proposed to extract image features. To improve the Convolution Restricted Boltzmann Machine (CRBM), a new training objective function to maximize the probability of intermediate image area is proposed, along with introducing the cross-entropy penalty factor and dropout training. After that, CDBM is designed based on modified CRBM. The mean-pool mechanism is presented to lessen computational complexity and improve the robustness of features for image scaling. The relationship between layers is simplified to extract high-level abstract features. The MNIST handwritten digits database is used to test this new model and the results prove that features extracted by CDBM are more accurate than CDBN. The classification accuracy rate increase at least 0.5%, and training time decrease more than 50%.

Keywords: deep learning; image features extraction; CRBM; CDBM

基于能量模型的受限玻尔兹曼机 RBM (restricted boltzmann machine)^[1] 以其简单的人工神经网络形式和快速的学习算法,已经广泛应用于数据降维、语音识别和图像处理等多个机器学习领域,进而催生出机器学习一个新的方向——深度学习^[2]. 目前以受限玻尔兹曼机为基础的图像处理方法通常采用两种方式构建模型:一是直接将图像中每一个像素对应于一个可见单元^[3];二是采用向量

化的多种特征作为可见单元^[4]. 这两种方式的弊端在于:模型只能处理较小的图像,难以处理大尺度图像;选取的是显式特征,受个人经验影响较大,灵活性差. 因此文献[5]提出了卷积受限玻尔兹曼机 CRBM(convolutional RBM),构建了卷积深度信念网络 CDBN(convolutional deep belief net)用于提取不同层次的图像表达,在手写字识别和人脸识别中取得了最好结果,但仍然存在计算复杂度高、训练缓慢等问题.

参考目前性能较好的其他神经模型^[6-8],本文提出一种基于改进 CRBM 的图像特征深度提取方法. 首先对 CRBM 进行改进,设计了图像的补零操作以及 CRBM 新训练目标,引入稀疏性能较好的交

收稿日期: 2005-04-01.

基金项目: 国家自然科学基金项目(61032001).

作者简介: 刘 凯(1986—),男,博士研究生;
张立民(1976—),男,教授,博士生导师.

通信作者: 张立民, wendao_2008@163.com.

叉熵稀疏惩罚因子抑制 CRBM 的特征同质化问题;然后提出了均值聚合机制和简化层间关系的深度学习模型——卷积深度玻尔兹曼机 CDBM (convolutional deep boltzmann machine).

1 卷积受限玻尔兹曼机

1.1 CRBM 模型

CRBM 是类似于 RBM 的两层结构. 设 CRBM 可见单元层为 $N_v \times N_v$ 的二值矩阵 V , 且包含 K 个 $N_w \times N_w$ 的卷积核, 则隐单元层由 K 个大小为 $N_h \times N_h$ 的特征映射面组成, 标记其中第 k 个特征映射面 H^k 的卷积核为 W^k , 偏置为 b_k . 该特征映射面中第 i 行第 j 列的隐单元条件激活概率为

$$P(h_{ij}^k = 1 | v) = \sigma((W^k * V)_{ij} + b_k).$$

式中: σ 为 sigmoid 函数, $(W^k * V)_{ij}$ 表示卷积核 W^k 与可见层 V 中以第 i 行第 j 列单元为左上角的大小为 $N_w \times N_w$ 矩形块卷积.

同理, 设定可见单元共享偏置为 c , 则可见单元层中单元 v_{ij} 的条件激活概率为

$$P(v_{ij} = 1 | h) = \sigma((\sum_k W^k * h^k)_{ij} + c). \quad (1)$$

参考文献[5], CRBM 的卷积操作如图 1 所示, 其中可见单元层中无阴影部分为图像的中间区域 V^m (大小为 $(N_v - 2(N_w - 1)) \times (N_v - 2(N_w - 1))$).

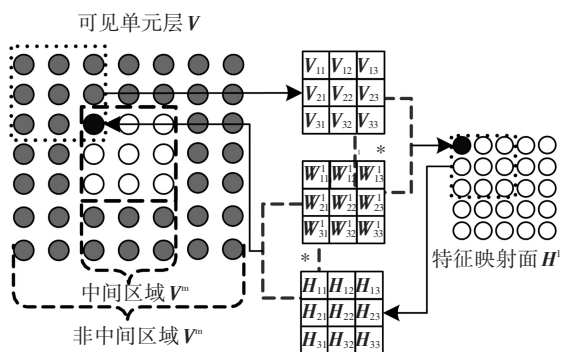


图 1 CRBM 卷积操作

从图 1 可以看出, 根据式(1)仅仅能够得到 V^m 区域单元的后验激活概率. 对于可见单元层的非中间区域 V^b , 由于其可见单元参与卷积的次数与 V^m 不同, 因此需要单独对待, 且文献[5]忽视了这一问题.

1.2 卷积受限玻尔兹曼机改进

针对 CRBM 模型可见单元重构问题, 提出 CRBM 的 3 个改进措施.

1.2.1 交叉熵稀疏惩罚因子

在 RBM 对网络连接权值进行学习的过程中, 存在的主要问题是特征学习的同质化. 由于所有隐单元都是在相互独立的基础上对训练数据进行学习, 当训

练数据中存在某种共有特征时, 隐单元均会受到影响, 表现在 RBM 学习到的连接权值 (即数据特征模式) 之间的列 (对应于每个隐单元) 相似度过高, 而解决这一问题的方法就是调节隐单元的稀疏性.

文献[9]通过对 RBM 训练的最大似然目标函数增加稀疏惩罚因子以抑制隐单元的激活概率. 但该方法存在的问题是: 惩罚因子只对隐单元偏置有影响, 且惩罚因子对于隐单元的稀疏化力度随着隐单元平均激活概率的下降而减弱. 文献[10]利用交叉熵的概念提出了基于交叉熵的稀疏惩罚因子, 取得了优于误差平方和的结果. 因此, 提出将交叉熵稀疏惩罚因子引入到 CRBM 训练, 即

$$P_E = \sum_{k=1}^K \left(p \log \left(\frac{1}{N} \sum_{n=1}^N q_k^n \right) + (1-p) \log \left(1 - \frac{1}{N} \sum_{n=1}^N q_k^n \right) \right).$$

式中: p 为稀疏度系数; N 为训练样本个数; K 为 CRBM 中卷积核个数; q_k^n 代表第 n 个样本下的卷积核 k 对应的特征映射面平均激活概率, 即

$$\frac{1}{N_h \times N_h} \sum_{i=1}^{N_h} \sum_{j=1}^{N_h} P(h_{ij}^k | v).$$

1.2.2 改进训练目标

鉴于 CRBM 可见单元重构的分析, 提出通过补零操作将边缘区域 V^b 纳入到中间区域得到 V^m' , 从而满足式(1)计算. 因此改进后的模型训练目标 F_{Target} 由文献[5]的最大化图像似然概率变为最大化补零后图像中间区域似然概率, 即

$$F_{\text{Target}} = \max \left(\sum_v \log P(V^m' | V^b) - P_E \right).$$

式中: V^m' 表示新图像中间区域重构数值, 尺寸为原始图像大小, 避免了文献[5]中图像非中间区域 V^b 不能通过式(1)直接计算得到的缺陷; V^b' 代表新图像的非中间部分, 由 0 组成; $V^m' | V^b'$ 表示 V^b' 保持不变; $-P_E$ 代表最小化交叉熵稀疏惩罚因子, 即使 CRBM 在学习过程中的特征映射面平均激活概率与 p 之间的 Kullback-Leiber 距离最小.

1.2.3 dropout 训练方法

引入 dropout 的目的在于防止训练过拟合, 其方式为在模型训练时以概率 q 随机设置网络中若干隐单元权值为 0.

2 卷积深度玻尔兹曼机

由于单层 CRBM 提取图像特征能力有限, 并为更好形成图像高层特征表达, 构建了基于 CRBM 的深度学习模型——卷积深度玻尔兹曼机 (CDBM). CDBM 较 CDBN 的优势在于计算复杂度较小且具有精准的聚合机制以及简洁的层间关联.

2.1 均值聚合机制

为使提取出的特征对输入样本具有较高的畸变

容忍能力,并增强算法对图像的缩放大小的健壮性,借鉴 CNN^[11],对 CDBM 中增加隐单元聚合机制。

文献[5]提出概率聚合(probabilistic pool)机制,目的在于适应 RBM 的双向连接结构,但问题是隐单元的聚合计算复杂度较高。文献[12]提出极大值聚合(Max Pool)机制,将 block 中隐单元的后验激活概率最大值作为聚合层对应点数值,其优势在于有效降低聚合机制的算法复杂度,但缺陷为易忽略图像中的细微特征,而这些特征对于图像识别往往是有有效的。

鉴于以上两种聚合机制的优缺点,提出新的聚合机制——均值聚合:将聚合层单元的值定义为对应 block 中所有隐单元后验激活概率的均值,即

$$p_{\alpha\beta}^k = \frac{\sum_{i=1+C\times(\alpha-1)}^{C,C} \sum_{j=1+C\times(\beta-1)}^{C,C} P(h_{ij}^k = 1 | V)}{C \times C} \quad (2)$$

式中: $p_{\alpha\beta}^k$ 表示聚合层中第 α 行第 β 列的单元值, $C \times C$ 为聚合层中 block 的尺寸。

从式(2)可以看出,均值聚合的计算复杂度远低于概率聚合。为直观显示不同聚合机制差异,以 Lena 图像为例,生成的聚合图像如图 2 所示,其中 block 大小为 4×4 。



图2 不同聚合机制下的聚合图像

从图 2 可以看出,极大值聚合下的图像细节最模糊;而概率聚合和均值聚合图像较为贴近原始图像。采用基于 SIFT 特征的图像匹配,3 幅聚合图像与原始图像之间的特征点匹配关系如表 1 所示。

聚合机制	原始图像	概率聚合	极大值聚合	均值聚合
概率聚合	34	65	34	52
极大值聚合	26	34	63	40
均值聚合	32	52	40	62

从表 1 数据可以看出,极大值聚合图像与原始图像匹配的特征点数最少,概率聚合与均值聚合数目接近,且概率聚合图像与均值聚合图像更为接近。结合图 2 可证明,采用概率聚合和均值聚合机制生成的聚合层图像能够保留原始图像较多特征,同时鉴于均值聚合较小的计算复杂度,使用均值聚合机制更具备优势。

2.2 CDBM 结构

参照 CDBN 结构,对 CDBM 模型层间关联进行设计。以 6 层 CDBM 为例,其模型结构如图 3 所示。模型输入层即 L1 层为原始图像本身;L2 层为低层特征提取层,与 L1 构成低层 CRBM;L3 层为聚合层,由 L2 层特征映射面通过均值聚合机制并相互叠加生成;L3 层与 L4 层构成高层 CRBM,完成图像底层特征向高层特征的组合;L5 层为 L4 的聚合合并层,并作为图像特征用于后续处理。与 CDBN 不同之处在于: CDBN 中 L3 层由多个聚合面组成,其二级 CRBM 卷积核为三维;CDBM 中 L3 由 L2 生成的各个聚合面叠加生成,二级 CRBM 的卷积核为二维。

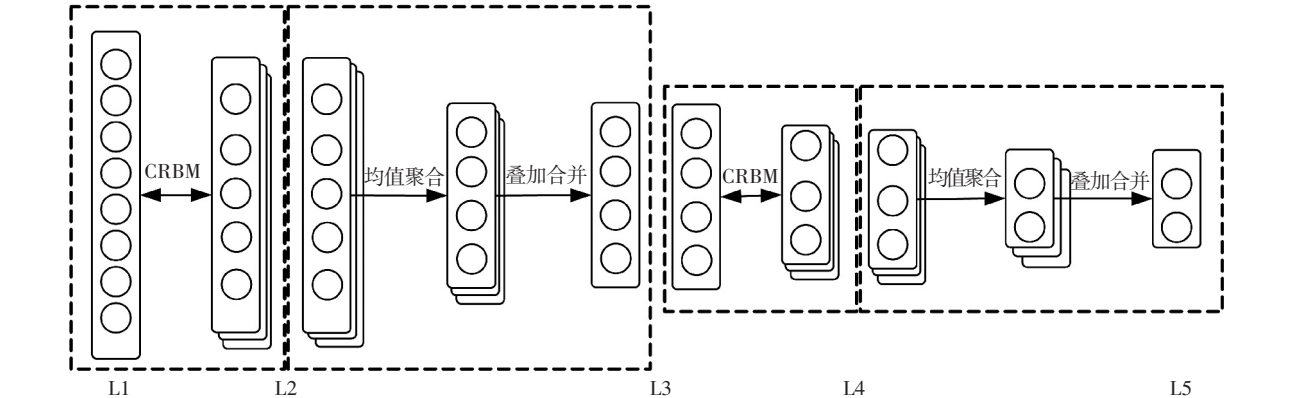


图3 CDBM 结构图

2.3 CDBM 训练

由于 CDBM 中 L2-L3 与 L4-L5 间为单向连接,因此 CDBM 的训练按照逐层贪婪无监督训练^[13]。即首先对每一级 CRBM 进行无监督学习,待 CRBM 达到数据收敛条件或超过训练迭代次数后,固定本层 CRBM 参数,随后继续下级 CRBM 训练。

3 实验

MNIST 数据集为 0 到 9 的 10 个大小为 28×28 手写数字图像集合,其中训练集个数为 60 000,测试集个数为 10 000^[14]。为加速训练和学习,将训练集的每 100 幅作为一个 batch,用于参数更新。实验平台采

用安装 Matlab(2013a)的主频为 2.4 GHz 的台式机.

3.1 CRBM 训练目标实验

本实验采用重构误差作为评价 RBM 模型训练效果标准,其值是以训练数据作为初始状态,计算经过若干次 Gibbs Sample 后与原数据的一阶范数差值. 为验证改进后新训练目标的学习,设置了如下 2 种训练过程(模型卷积核个数为 24,大小为 7×7 ,隐单元稀疏度系数 $p = 0.05$,学习速率 $\eta = 0.01$,循环迭代次数为 20).

- 1)原始训练过程:以最大化图像重构概率为训练目标,选用误差平方和稀疏惩罚因子.
 - 2)新训练过程:以最大化补零图像重构概率为训练目标,选用误差平方和稀疏惩罚因子.
- 2 种训练方法过程中的重构误差如图 4 所示.

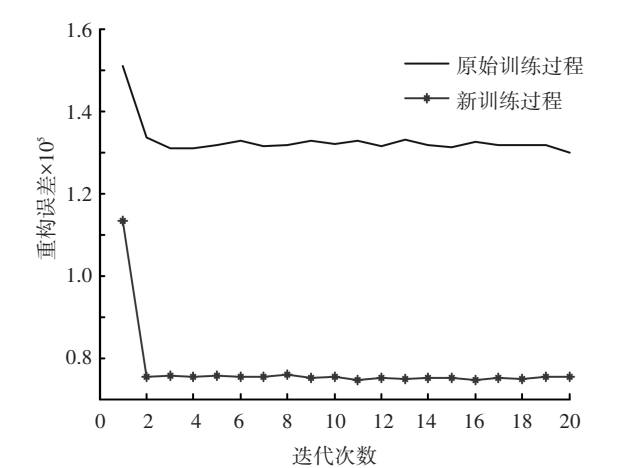


图 4 两种训练方法的重构误差对比

从图 4 可以看出,新训练过程的模型重构误差较低,表明经训练后的 CRBM 对数据具有更好的似然度. 其原因在于新训练过程完成了对原始图像边缘区域的重构,从而降低了模型训练过程中的重构误差.

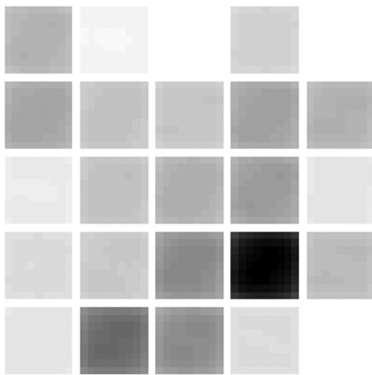
3.2 稀疏惩罚因子实验

为验证不同稀疏惩罚因子对 CRBM 训练的影响,结合实验 1 结果,选取新的训练目标函数和相同的模型设置进行 CRBM 训练. 图 5 为在 3 种不同稀疏惩罚因子下的 CRBM 卷积核可视化图.

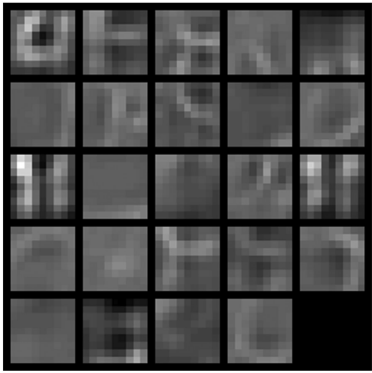
从图 5 可以看出:在没有稀疏惩罚因子的情况下,CRBM 出现了严重的特征同质化现象,绝大多数卷积核相似并且没有对应的特征表现;采用稀疏惩罚因子后,卷积核之间差异变大且都有较为明显的数字笔画部分,证明了稀疏惩罚因子能够有效地避免特征同质化问题;从图 5(b)和图 5(c)对比发现,图 5(b)中仍然存在着个别卷积核相似的现象,图 5(c)的卷积核之间差异变大,说明相较于误差平方和稀疏惩罚因子,在交叉熵稀疏惩罚因子下 CRBM 学习的特征更局部化,克服 CRBM 特征同质化的效

果较好,证明特征同质化得到进一步弱化.

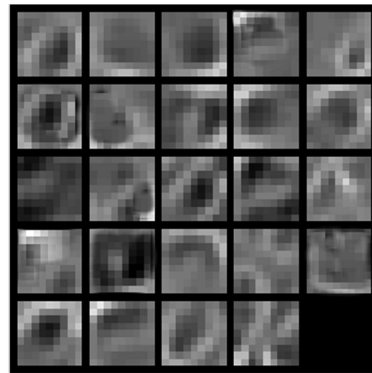
完成 CRBM 训练后,选用特征映射面作为手写字符特征,采用 LIBSVM^[15] 提供的径向基支持向量机(RBF-SVM)和线性支持向量机(LSVM)作为最终分类器,其中参数设置除了核函数选择不同外,其余均为默认设置,其实验结果如表 2 所示.



(a) 无稀疏惩罚因子



(b) 误差平方和稀疏惩罚因子



(c) 交叉熵稀疏惩罚因子

图 5 稀疏惩罚因子对于 CRBM 卷积核影响

表 2 MNIST 分类准确率

稀疏惩罚机制	特征向量维数	分类准确率	
		RBF	LSVM
无	122	0.12	0.20
误差平方和	122	97.60	95.60
交叉熵	122	98.30	95.65

从表 2 数据可以看出,采用交叉熵稀疏惩罚因子的 CRBM 分类准确率最高,而没有采用稀疏惩罚机制的 CRBM 训练无效,且选用 RBF-SVM 为分类器的准确率要高于 LSVM,其原因在于 CRBM 中的可见单元通常参与多次卷积运算,这会增强模型训练时原始数据共有特征的影响,因此选用稀疏惩罚较好的交叉熵因子能够改善特征同质化带来的模型特征分辨力差问题.

4 结 论

CDBM 实现了从图像底层特征到高层特征的提取,其过程符合实际的生物神经网络,并且实验结果证明了该模型能够取得良好效果. CDBM 通过训练数据进行无监督学习,避免了以往显式的特征抽取,而是对训练数据进行隐式学习,减少了个人经验对图像特征提取的影响. 相比于 CRBM 构建的深度网络 CDBN,经优化后的 CDBM 训练时间更短,并可以有效完成图像特征的提取. 今后工作应继续改进 CRBM 学习算法过程,减少模型训练时间,详细分析模型参数对于训练学习的影响,并将新模型应用到更多领域.

参考文献

[1] 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.

[2] 林妙真. 基于深度学习的人脸识别研究[D]. 大连: 大连理工大学, 2013.

[3] LAROCHELLE H, ERHAN D, COURVILLE A, et al. An empirical evaluation of deep architectures on problems with many factors of variation [C]// Proceedings of the 24th international conference on Machine learning. New York: ACM, 2007: 473-480.

[4] SRIVASTAVA N, SALAKHUTDINOV R. Multimodal learning with deep boltzmann machines[C]// Advances in neural information processing systems 2012. New York: ACM, 2012: 2222-2230.

[5] LEE H, GROSSE R, RANGANATH R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations[C]//

Proceedings of the 26th Annual International Conference on Machine Learning. New York: ACM, 2009: 609-616.

[6] GOODFELLOW I, WARDE-FARLEY D, MIRZA M, et al. Maxout Networks [C]// Proceedings of the 30th International Conference on Machine Learning. Piscataway, NJ: IEEE, 2013: 1319-1327.

[7] CIRESAN D C, MEIER U, MASCI J, et al. Flexible, high performance convolutional neural networks for image classification [C]// Proceedings International Joint Conference on Artificial Intelligence 2011. New York: ACM, 2011: 22(1) . 1237-1239.

[8] CIRESAN D, MEIER U, SCHMIDHUBER J. Multi-column deep neural networks for image classification[C]// IEEE Conference on Computer Vision and Pattern Recognition 2012. Piscataway, NJ: IEEE, 2012: 3642-3649.

[9] LEE H, EKANADHAM C, NG A Y. Sparse deep belief net model for visual area V2 [C]// Advances in Neural Information Processing Systems 2008 . New York: ACM, 2008: 873-880.

[10] HINTON G. A practical guide to training restricted boltzmann machines [M]. Berlin: Springer Berlin Heidelberg, 2012: 599-619.

[11] LAWRENCE S, GILES C L, TSOI A C, et al. Face recognition: A convolutional neural-network approach [J]. IEEE Transactions on Neural Networks, 1997, 8(1): 98-113.

[12] NOROUZI M. Convolutional restricted boltzmann machines for feature learning [D]. Vancouver: School of Computing Science-Simon Fraser University, 2009.

[13] LE Q V. Building high-level features using large scale unsupervised learning [C]// Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE, 2013: 8595-8598.

[14] LIU C L, NAKASHIMA K, SAKO H. Handwritten digit recognition: benchmarking of state-of-the-art techniques [J]. Pattern Recognition, 2003, 36(10): 2271-2285.

[15] CHANG C C, LIN C J. LIBSVM: a library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 27-31.

(编辑 王小唯 苗秀芝)