

doi:10.16055/j.issn.1672-058X.2015.0008.011

基于深度学习的手写数字分类问题研究*

宋志坚¹, 余 锐²

(1.重庆交通大学 信息科学与工程学院,重庆 400074;2.重庆交通大学 交通运输学院,重庆 400074)

摘 要:手写体数字因其书写风格差异大、上下文无关及识别准确度要求高等原因导致其识别难度大,针对手写体数字识别的特点及要求,使用深度学习算法进行分类,通过对样本的训练完成手写体数字的识别,同时与 SVM 算法及 BP 神经网络分类效果进行对比;实验结果表明深度学习在识别手写体数字时具有更高的准确率。

关键词:手写体数字识别;深度学习;样本训练

中图分类号: O224 **文献标志码:** A **文章编号:** 1672-058X(2015)08-0049-05

由于不同的人所写的数字体形态各异,千差万别,书写不规范导致手写数字识别的复杂性,而手写体数字的识别是建立在特征提取及比较的基础之上的。按照提取字符特征的不同,现有的数字手写体识别算法大体上可以分为两类:一类是基于结构特征的手写体数字识别算法。它们通过识别字符图像内部包含的凹陷区特征、轮廓特征结构突变点特征等基元,采用模板匹配的方式实现手写体数字的自动识别^[1,2]。这类方法能够直观地描述字符的结构,但是存在着对字符形变及噪声缺乏鲁棒的问题;另外一类是基于统计特征的手写体数字识别算法。这类算法基于对大量样本的表征、变换和学习,通过估计不同样本类别的特征空间分布训练相应的分类器,并利用这些分类器对未知模式进行分类^[3,4]。当训练样本选取得足够充分时,这类方法能够具有很好的识别能力。

针对手写体识别的特点及要求,分析现有算法存在的问题,使用深度学习进行手写体数字识别。算法能够通过使用大量的简单神经元组成的网络,利用输入与输出之间的非线性关系,对复杂函数进行近似,对观测样本进行拟合,并在学习输入样本本质特征的抽取上体现了强大的能力,实现了手写体数字的自动识别。

1 深度学习

深度学习(Deep Learning)的概念由 Geoffrey Hinton 等人提出,与传统机器学习相比,整合特征抽取和分类器到一个学习框架下,克服了传统方法中,特征选取困难等问题。

深度学习是通过大量的简单神经元组成,每层的神经元接收更低层的神经元的输入,通过输入与输出之间的非线性关系,将低层特征组合成更高层的抽象表示,并发现观测数据的分布式特征。通过自下而上的学习形成多层的抽象表示,并多层次的特征学习是一个自动地无人工干预的过程。根据学习到的网络结

收稿日期:2014-10-21;修回日期:2014-12-11.

* 基金项目:重庆市自然科学基金计划项目(CSTC2011jjA0893).

作者简介:宋志坚(1990-),男,内蒙古包头人,硕士研究生,从事机器学习研究.

构,系统将输入的样本数据映射到各种层次的特征,并利用分类器或者匹配算法对顶层的输出单元进行分类识别等。

1.1 深度学习的基本思想

假设一个系统 S ,它是一个 n 层(S_1, S_2, \dots, S_n)的结构, I 是系统的输入, O 是系统输出,形象地表示为: $I \Rightarrow S_1 \Rightarrow S_2 \Rightarrow \dots \Rightarrow S_n \Rightarrow O$,如果输出 O 等于输入 I ,物理意义也就是表明在经过系统变化之后,输入 I 的信息量没有任何损失,和原始的输入保持了不变,这表明了输入 I 经过每一层 S_i 均没有丢失任何信息,即在任一层 S_i ,它都是输入 I 也就是原始信息的另外一种表示。

1.2 受限玻尔兹曼机(RBM)

1.2.1 RBM 的模型定义

与传统神经网络相比,深度学习一个重要的突破就在于,在一定程度上克服了浅层神经网络训练的效率和效果问题。而在 Hinton 等人提出来的框架中,这种深度学习的学习策略,就是在进行多层的模型全局学习之前,先将多层神经网络分解为若干个受限玻尔兹曼机(Restricted Boltzmann Machine,简称 RBM)的叠加,并逐层训练 RBM。因此 RBM 作为深度学习中一个非常重要的基础模型。

RBM 是一个随机的二值化对称连接的神经网络,是一种特殊形式的对数线性马尔可夫随机场(MRF),也就是说其能量函数关于其自由参数是线性的。它包含了一组可视单元 $v \in \{0, 1\}^D$, 和一组隐藏单元 $h \in \{0, 1\}^F$ 。当参数一定时,对于一个受限的玻尔兹曼机(RBM)来说,其能量定义为

$$\begin{aligned} E(v, h; \theta) &= -v^T W h - b^T v - a^T h \\ &= \sum_{i=1}^D \sum_{j=1}^F W_{ij} v_i h_j - \sum_{i=1}^D b_i v_i - \sum_{j=1}^F a_j h_j \end{aligned} \quad (1)$$

这里 $\theta = \{W, b, a\}$ 是 RBM 模型的参数,其中 W_{ij} 表示可视层节点 v_i 到隐层节点 h_j 的链接权重, b_i 和 a_j 是偏置项。RBM 的状态符合玻尔兹曼分布的形式,即可视节点和隐层节点的联合概率分布被定义为

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

$$Z(\theta) = \sum_v \sum_h \exp(-E(v, h; \theta)) \quad (3)$$

其中, $P(v, h; \theta)$ 这种分布被称之为玻尔兹曼分布函数;而 $Z(\theta)$ 是一个分布和,又或者叫做归一化常数。那么 RBM 模型分配给可视节点 v 的概率:

$$P(v; \theta) = \frac{1}{Z(\theta)} \sum_h \exp(-E(v, h; \theta)) \quad (4)$$

由于 RBM 模型二分结构的这一特殊性,隐层节点可以清楚地同可视层区分开来:

$$\begin{aligned} P(v; \theta) &= \frac{1}{Z(\theta)} \sum_h \exp(v^T W h + b^T v + a^T h) \\ &= \frac{1}{Z(\theta)} \exp(b^T v) \prod_{j=1}^F \sum_{h_j \in \{0, 1\}} \exp(a_j h_j + \sum_{i=1}^D W_{ij} v_i h_j) \\ &= \frac{1}{Z(\theta)} \exp(b^T v) \prod_{j=1}^F (1 + \exp(a_j + \sum_{i=1}^D W_{ij} v_i)) \end{aligned} \quad (5)$$

由于在 RBM 中,同一层节点之间是互不相联的,因此,在给定其中一层节点状态时,另一层节点之间的状态条件分布相互独立,即:

$$P(h | v; \theta) = \prod_j p(h_j | v) \quad (6)$$

$$P(v | h; \theta) = \prod_i p(v_i | h) \quad (7)$$

$$P(h_j = 1 | v) = g\left(\sum_i W_{ij}v_i + a_j\right) \quad (8)$$

$$P(v_i = 1 | h) = g\left(\sum_j W_{ij}h_j + b_i\right) \quad (9)$$

其中, $g(x)$ 为 sigmoid 函数, $g(x) = 1/(1 + \exp(-x))$ 是一个逻辑函数。

1.2.2 RBM 的学习

RBM 可利用极大似然法学习^[5]。在给定训练数据 v_1, v_2, \dots, v_D 。学习 RBM 相当于最大化目标函数 $\log P(v; \theta)$ 。这种情况下, 先对一般数据进行推导。给定某个点 x 可以用函数的形式表示记为 $f(x; \theta)$, 此处 θ 是模型的参数向量。那么 x 的概率 $P(x; \theta)$ 可以表示为

$$p(x; \theta) = \frac{1}{Z(\theta)} f(x; \theta) \quad (10)$$

其中, $Z(\theta)$ 是分配函数, 定义为

$$Z(\theta) = \int f(x; \theta) dx \quad (11)$$

学习模型参数 θ , 可以通过最大化一组训练数据 $X = x_1, x_2, \dots, x_k$ 的概率, 即:

$$p(X; \theta) = \frac{1}{K} \prod_{k=1}^K \frac{1}{Z(\theta)} f(x_k; \theta) \quad (12)$$

那么:

$$\begin{aligned} L &= \frac{\partial \log P(X; \theta)}{\partial \theta} = \frac{1}{K} \sum_{k=1}^K \frac{\partial \log f(x_k; \theta)}{\partial \theta} - \frac{\partial \log Z(\theta)}{\partial \theta} = \\ &= \left[\frac{\partial \log f(x; \theta)}{\partial \theta} \right]_x - \frac{1}{Z(\theta)} \frac{\partial Z(\theta)}{\partial \theta} = \\ &= \left[\frac{\partial \log f(x; \theta)}{\partial \theta} \right]_x - \frac{1}{Z(\theta)} \int \frac{\partial f(x; \theta)}{\partial \theta} dx = \\ &= \left[\frac{\partial \log f(x; \theta)}{\partial \theta} \right]_x - \frac{1}{Z(\theta)} \int f(x; \theta) \frac{\partial \log f(x; \theta)}{\partial \theta} dx = \\ &= \left[\frac{\partial \log f(x; \theta)}{\partial \theta} \right]_x - \int \frac{f(x; \theta)}{Z(\theta)} \frac{\partial \log f(x; \theta)}{\partial \theta} dx = \\ &= \left[\frac{\partial \log f(x; \theta)}{\partial \theta} \right]_x - \int P(x; \theta) \frac{\partial \log f(x; \theta)}{\partial \theta} dx = \\ &= \left[\frac{\partial \log f(x; \theta)}{\partial \theta} \right]_x - \left[\frac{\partial \log f(x; \theta)}{\partial \theta} \right]_{P(x; \theta)} \end{aligned} \quad (13)$$

可根据减号将公式分为正项 (positive phase) 和负项 (negative phase) 两个部分。对于正项部分, 表示的是满足训练数据分布中的数据期望, 可以通过遍历所有训练数据算出来。而对于负项部分, 表示的是概率密度分布为 $p(x; \theta)$ 的函数 $f(x)$ 的均值。

同样地, RBM 模型的参数以及其分布概率函数, 根据极大似然准则可以从公式 (1) 推导出:

$$\frac{\partial \log P(v; \theta)}{\partial W} = \alpha (E_{\text{data}}[vh^T] - E_{\text{model}}[vh^T]) \quad (14)$$

$$\frac{\partial \log P(v; \theta)}{\partial a} = \alpha (E_{\text{data}}[h] - E_{\text{model}}[h]) \quad (15)$$

$$\frac{\partial \log P(v; \theta)}{\partial b} = \alpha (E_{\text{data}}[v] - E_{\text{model}}[v]) \quad (16)$$

在 RBM 中, 正项表示满足 $P_{\text{data}}(h, v; \theta) = P(h|v; \theta) P_{\text{data}}(v)$ 这数据分布的平均值, 而负项是满足模型定义的分布的数据期望, 这个均值在采用极大似然算法中复杂度为 $\min\{D, F\}$ 的指数级, 也就是输入节点或隐藏层

节点数目的指数级。这个积分的计算通常来讲不是线性可算的。此时对分布 $p(x)$ 下的样本进行大量采样并求其均值就可以逼近原积分结果。对于各态遍历的马尔可夫过程,随着转移次数的增多,随机变量的取值分布将不受随机变量初始值的影响,最终都会收敛于唯一的平稳分布^[6]。这意味着,不论随机变量的初始值设为多少,经过足够多次转移之后,变量取各值的概率总会不断接近于该过程的平稳分布。显然,如果为了获取某个目标分布下采样,只需要模拟以其为平稳分布的马尔可夫过程,并执行足够多次的概率转移得到一个与样本分布接近的平衡分布。

1.3 深度信念网络

全局优化具有多个隐藏层的深度信赖网络往往是难以处理的,而为了取得较好的优化性能,贪婪算法是可以用于此处的,即逐层优化,每次只学习相邻两层 RBM 模型参数,并通过这样子贪婪地逐层学习以获得全局的 DBNs^[7]。通过贪婪算法获得的深度信赖网络根据最终的感兴趣的判别准则进行微调就可以获得最终深度信赖网络,也就是把一个 DBNs 网络分层,对每一层进行无监督学习,最后对整个网络用监督学习进行微调。过程可以归为两部分:

(1) 预训练。深度信赖网络分解成由相邻两层构成的一系列受限的玻尔兹曼机,逐层训练参数,初始化深度信赖网络:首先将经验数据 v 作为输入,训练第一层受限玻尔兹曼机的权值系数矩阵 W_1 ;接着将 W_1 固定,通过 $p(h_1|v) = p(h_1|v, W_1)$,训练出第一层受限玻尔兹曼机的隐层向量 h_1 ;将 h_1 作为第二层受限玻尔兹曼机的输入,训练第二层受限玻尔兹曼机的权值系数矩阵 W_2 ;递归地计算出每一层的隐含单元向量和权值系数矩阵。

(2) 微调。为了使得模型具有更好的表现能力,当完成一组 RBM 之后,将其展开,然后采用 back propagation 的方法,运用梯度下降法对整个 DBNs 进行优化。

这个基于贪婪的学习方法无论是在预训练还是微调阶段在时间复杂度上和空间复杂度上均是线性的。所以在进行大规模数据的学习时,这个方法是非常有效的。

2 实例结果及结论

在实验中选取样本库中手写体数字进行训练及测试,实现分类预测,表 1 为选取样本数。分别根据每个数字的训练集进行训练,使用测试集测试效果,对于每种数字识别正确率如表 2。

表 1 训练样本数及测试样本数

	数字 0	数字 1	数字 2	数字 3	数字 4	数字 5	数字 6	数字 7	数字 8	数字 9
训练集	5 900	6 700	5 900	6 100	5 800	5 400	5 900	6 200	5 800	5 900
测试集	980	1 100	1 000	1 000	980	890	950	1 000	974	1 000

表 2 实验识别结果统计

											%
	0	1	2	3	4	5	6	7	8	9	
Deep Learning 算法	98	98.6	98.1	98.7	98.5	98.2	98	98.3	98.5	98.2	
SVM 算法	97	98	96.4	98	98	97	96.8	96.8	97.3	97	
BP 神经网络算法	91	93	92	91	91	93	91	92	91	93	

从实验结果可看出,SVM 算法优于 BP 神经网络算法,识别效果大幅度提高,而 Deep Learning 算法又在 SVM 算法的基础进一步提高,达到 98% 以上,识别效果令人满意。

3 结 语

深度学习算法的优越性逐步在被发掘,在数字识别领域已经达到一个高度,同时其推广程度相当高,还可以使用在手写汉字系统、图片搜索引擎等方面具有巨大的发展潜力。

参考文献:

- [1] 陈军胜. 组合结构特征的自由手写体数字识别算法研究[J]. 计算机工程与应用, 2013, 49(5): 179-184
- [2] 吴忠, 朱国龙, 黄葛峰, 等. 基于图像识别技术的手写数字识别方法[J]. 工业控制计算机, 2011, 21(12): 48-51
- [3] 石会芳, 胡小兵, 刘瑞杰, 等. 基于启发式 GA-SVM 的手写数字字符识别的研究[J]. 计算机技术与发展, 2012, 22(10): 5-9
- [4] 石会芳. 支持向量机及其在手写数字识别中的应用[D]. 重庆: 重庆大学, 2013
- [5] HINTON G. A Practical Guide to Training Restricted Boltzmann Machines[J]. Report of Momentum, 2010, 9(1): 1-20
- [6] LIU J S. Monte Carlo Strategies in Scientific Computing[M]. Berlin Heidelberg: Springer Verlag, 2008
- [7] BENGIO Y, LAMBLIN P, POPOVICI D, et al. Greedy Layer-wise Training of Deep Networks [J]. Advances in Neural Information Processing Systems, 2007, 19-153

Research on the Classification of Handwriting Number Based on Deep Learning

SONG Zhi-jian¹, YU Rui²

(1.School of Computer science and technology, Chongqing Jiaotong University, Chongqing 400074, China; 2.School of transportation, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: It is difficult to recognize handwriting number, because writing styles are different, context is free and recognition requires high accuracy. According to the characteristics and requirements of handwriting number recognition, this paper uses the Deep Learning algorithm for classification, based on the sample of training to recognize the handwriting number.

At the same time, the experiment compared the classification effect with that of SVM algorithm and BP neural network. The experimental results show that the Deep Learning of handwriting number in recognition has a higher accuracy.

Keywords: handwriting number recognition; Deep Learning; sample training