

基于狼群算法的 DBN 网络结构确定方法研究

周贤泉, 宋 威

(江南大学 物联网工程学院, 江苏 无锡 214122)

摘 要: 由于深度信念网络(DBN)结构难以确定,提出运用狼群算法确定 DBN 每层神经元个数, K-means 聚类准确性确定是否增加隐含层的方法确定 DBN 结构. 根据狼群算法以最小化所有样本重构误差的平方和为目标函数来确定 DBN 每层神经元的个数,以确定 DBN 的初步结构. 为了检验 DBN 结构的有效性,利用 DBN 提取的数据特征进行聚类测试,进一步根据聚类结果来修正 DBN,以获取 DBN 的最佳结构. 实验中选取 Iris 数据集进行聚类测试,其结果表明,通过所提出的方法获取的 DBN 有效结构来处理原始数据,能够提高聚类的准确性.

关键词: 深度信念网络;狼群算法;重构误差;聚类测试

中图分类号: TP391.3

文献标识码: A

文章编号: 1000-7180(2018)07-0028-07

Research on DBN Network Structure Determination Method Based on Wolves Algorithm

ZHOU Xian-quan, SONG Wei

(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

Abstract: It's difficult to determine the structure of deep belief network (DBN), so raise the method that use the wolves algorithm to determine the number of DBN per layer of neurons, and K-means Clustering Accuracy Determines whether or not to increase the hidden layer method. According to the minimal reconstruction error function of wolves algorithm, it can work out the quantity of each neuron, so as to determine the initial structure of the DBN. In order to verify the effectiveness of the DBN structure, using DBN to extract the data characteristics of the clustering test, so can get the final structure of the DBN. The Iris dataset was tested in the experiment, the results show that the proposed method handles the raw data by the effective structure of the DBN obtained and improve the accuracy of clustering.

Key words: deep belief networks; wolves algorithm; reconstruction error; clustering test

1 引言

2006 年 Hinton 等人提出了深度信念网络(deep belief networks, DBN)^[1]模型,此模型能够从高维度的数据中提取到这些高维数据的有用特征数据,这些有用的特征数据往往更有助于应用研究.虽然 DBN 模型已经成功的应用于多个领域,但是现在仍然处于发展的初期,还有许多问题值得深入

研究.

从神经网络^[2]提出到现在,一直没有一个很好的方法来确定神经网络中隐含层节点的个数.在模式识别的研究中,如果隐含层的节点个数较少,那么网络的拟合能力与识别的能力就较差;如果隐含层的节点个数偏多,那么神经网络在训练的时候会浪费大量的时间,也会导致网络的识别能力较差.从神经网络提出到现在国内外很多学者对网络中隐含层

收稿日期: 2017-10-23; 修回日期: 2017-11-15

基金项目: 国家自然科学基金(61673193);中央高校基本科研业务费专项资金(JUSRP51635B, JUSRP51510);江苏省自然科学基金(BK20150159)

的节点数确定方法做了大量的研究,提出的方法主要有:递增法,试凑法,遗传算法等方法。递增法就是开始设置隐含层节点数为 1,然后逐渐增加隐含层节点的个数,最终找到最佳的网络结构;试凑法^[3-4]是不断的选择隐含层节点的个数,最终寻找到最佳的网络结构;遗传算法寻得最佳网络结构的方法通常是结合递增法和剪枝方法。比如:Benardos^[5]就是根据遗传算法与试凑法结合起来最终寻得最佳的网络结构。然而这些方法都是应用在普通的神经网络中,并没有应用在 DBN 网络模型中,所以本文根据文献^[5]的思想,提出一种新的 DBN 网络结构确定方法。

因 DBN 涉及的范围较广,难以给出一个标准方法来确定 DBN 模型结构,但是可以将确定 DBN 结构问题进行转换,即不是以人为的方式对 DBN 的宽度与深度进行试凑决定,而是将 DBN 结构与应用研究结合考虑来确定 DBN 合适的宽度与深度。所以本文提出运用狼群算法确定 DBN 每层神经元个数,K-means 聚类准确性确定是否增加隐含层的方法确定 DBN 网络的结构。

2 基于受限玻尔兹曼机(RBM)的无监督学习

2.1 受限玻尔兹曼机结构

受限玻尔兹曼机^[6](Restricted Boltzmann Machine,RBM)是一种特殊的玻尔兹曼机结构,可以看成是一种有向无环图。RBM 结构示意图如图 1 所示。

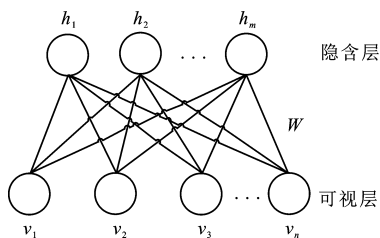


图 1 RBM 结构示意图

一般情况下,RBM 可以看成是构建 DBN 网络模型的基本单元。RBM 由隐含层单元 h 和可视层单元 v 组成,只有可视层与隐含层之间有连接。如图 1 所示,在一个受限玻尔兹曼机结构中,可视层单元 v 可以表示输入数据,可视层单元个数 n 表示输入数据的维度;隐含层数据可以用 h 来表示,其中 m 表示隐含层单元个数即提取特征数据的维度。RBM 可以

看成是一个基于能量的模型,隐含层与可视层之间 (v, h) 具有能量,

能量公式可以用式(1)来表示:

$$E(v, h) = - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m v_i W_{ij} h_j \quad (1)$$

式(1)中的 a_i 表示可视层的偏置, b_j 表示隐含层的偏置, W_{ij} 表示隐含层单元与可视层单元之间的权值。如果可视层单元已知,那么隐含层单元之间相互独立;如果隐含层单元已知,那么可视层单元之间也相互独立;隐含层与可视层的单元激活的概率一般情况下都是将 RBM 的能量函数先指数化,然后再归一化到 $p(v, h)$,如公式(2)所示:

$$p(v, h) = \frac{1}{Z} \sum_h e^{-E(v, h)} \quad (2)$$

Z 是一个归一化常数,可以看成是隐含层与可视层节点的所有激活概率的相加:

$$Z = \sum_{v, h} e^{-E(v, h)} \quad (3)$$

RBM 模型中,可视层向量的概率可以看成是将可视层向量对应的所有隐含层向量的概率值相加:

$$p(v) = \frac{1}{Z} \sum_h e^{-E(v, h)} \quad (4)$$

由于可视层单元之间,隐含层单元之间相互独立,所以有:

$$p(h | v) = \prod_{j=1}^m p(h_j | v) \quad (5)$$

$$p(v | h) = \prod_{i=1}^n p(v_i | h) \quad (6)$$

当可视层 v 状态固定时,隐含层单元 h_j 的二进制状态为 1 的概率为:

$$p(h_j = 1 | v) = \sigma(b_j + \sum_{i=1}^n v_i w_{ij}) \quad (7)$$

当隐含层 h 的状态时,可视层单元 v_i 的二进制状态为 1 的概率为:

$$p(v_i = 1 | h) = \sigma(a_i + \sum_{j=1}^m w_{ij} h_j) \quad (8)$$

式中,激活函数一般为: $\sigma(x) = \frac{1}{1 + e^{-x}}$ 。

RBM 模型具有两个主要特点:①模型推理比较简单,如果可视层向量已知,那么隐含层向量的分布可以看成是每个隐含层单元独立分布的乘积,根据公式(7)计算出隐含层单元的激活概率,从而得到后验分布采样。②将 RBM 模型堆叠起来,也能对堆叠

起来的深度网络进行逐层学习。

2.2 RBM 无监督训练

在 RBM 网络的训练过程中,文献[7-10]中提到的对比散度算法(Contrastive Divergence, CD)可以明显的减少 RBM 训练的时间. RBM 通过 CD 算法训练的主要思想是:将训练数据看成是原始的状态,将其从 RBM 可视层输入,然后根据文献[11]中的 Gibbs 采样,经过较少次数的状态转移,就可以提取到输入数据的主要特征. 以下六个步骤详细描述出 RBM 的训练以及权值的优化过程:

(1) 首先初始化 RBM 网络的一些参数,随机初始化隐含层的偏置向量 a 、可视层的偏置向量 b 、隐含层与可视层之间的权值矩阵 W .

(2) 输入数据从可视层层输入,根据公式(5)和公式(7)可以计算出隐含层单元的输出矩阵 h 的激活概率 $p(h | v)$. 输入数据前向传播的概率可以看成是可视层的输入矩阵 v 和隐含层的输出矩阵 h 的激活概率对应的乘积矩阵.

(3) 此时步骤(2)中得到的 $p(h | v)$ 为概率值,为了防止输出数据过度拟合,根据文献[6]可以将输出数据进行二值化处理.

(4) 首先对步骤(3)中处理得到的二值化值进行反向传播,再结合公式(6)、(8)计算出可视层的节点对应的重构特征矩阵 v' 的激活概率,那么就可以得到输入数据的重构特征数据.

(5) 然后再对重构的特征矩阵 v' 进行正向传播,根据步骤(2)可以计算出隐含层单元对应的激活概率矩阵 h' . 输入数据对应的重构特征 v' 和隐含层单元对应的激活概率矩阵 h' 的乘积可以看成是反向传播的概率.

(6) 隐含层 h 对应的偏置向量 b 的变化量可以根据步骤(2)中隐含层激活概率矩阵 h 与步骤(5)中得到的激活概率 h' 相减得到. 可视层 v 对应的偏置向量 a 的变化量可以根据可视层 v 的激活概率与重构的 v' 的激活概率相减得到. 隐含层与可视层之间的权值变化量可以根据步骤(2)中得到的正向传播概率向量与步骤(5)中得到的反向传播概率向量相减得到. 在 RBM 训练过程中,隐含层与可视层对应的权值与偏置都是同时更新的,所有偏置与权值是同时收敛的. 根据公式(9)、(10)、(11)再结合相

应的学习率,可以更新隐含层与可视层之间的权值于偏置:

$$W_{t+1} = W_t + \epsilon(v^* p(h | v) - v'^* p(h' | v')) \quad (9)$$

$$b_{t+1} = b_t + \epsilon(p(h | v) - p(h' | v')) \quad (10)$$

$$a_{t+1} = a_t + \epsilon(p(h | v) - p(h' | v')) \quad (11)$$

3 深度信念网络

3.1 深度信念网络结构

深度信念网络^[1](deep belief networks, DBN)是以受限玻尔兹曼机作为基本单元搭建的信任网络,由一个可视层与多个隐含层组成,是一种基于概率的生成模型,其结构如图 2 所示:

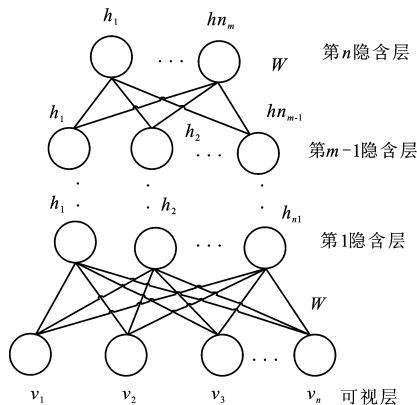


图 2 深度信念网络结构

3.2 深度信念网络训练

DBN 训练主要有两个步骤. 如图 2 所示,首先可以把输入层 n 个节点和第一隐含层中 n_1 个节点看成一个 RBM 网络,根据 1.2 节关于 RBM 网络的训练方法来训练第一个 RBM,可以得到第一个 RBM 结构对应的权值与偏置,根据对应的权值与偏置可以得到训练样本对应的隐含层的特征数据. 然后可以把第一个 RBM 得到的特征数据看成是第 2 个 RBM 网络可视层的输入值,将第二隐含层与第一隐含层看成是可视层为 n_1 个节点和隐含层为 n_2 个节点的 RBM 网络. 固定第一个 RBM 网络的权值和偏置,根据 1.2 节 RBM 网络训练方法可以计算出第二个 RBM 网络的输出值,偏置和权值. 以此类推训练 DBN 中接下来第三个,第四个 RBM 网络,直到训练完最后一个 RBM 网络,将得到一个无监督 DBN 模型. 通过 DBN 模型,输入数据从可视层输入,最后一个隐含层的输出数据即可以看成是所

提取的原始数据主要特征。

4 DBN 网络的结构确定方法

本文提出运用该算法以重构误差^[12]作为适应度函数(由于重构误差的计算比较简单),确定 DBN 每层神经元个数,然后以 K-means 聚类算法的准确性作为指标确定是否增加隐含层来改变 DBN 结构。

4.1 狼群算法智能行为和规则描述

(a)头狼的产生. 首先计算出各人工狼在初始的解空间中对应的函数值,取其中函数值最小的人工狼作为头狼. 在狼群算法迭代^[13]的过程中,每个人工狼都要与前一代的头狼对应的函数值作比较,如果此人工狼对应函数值比头狼所对应的函数值更小(为论述方便,以下以极小值问题进行讨论),则更新头狼位置,头狼不执行游走与召唤行为。

(b) 游走行为. 将头狼以外最优的 S_num 匹人工狼看成是探狼,在对应的解空间中寻找更好的猎物. 探狼 X_i 先自主决策,朝两个方向前进,如果朝某一个方向前进过程中搜索到更好的猎物 Y_i ,则比较与头狼对应函数值 Y_{lead} ,如果 $Y_i < Y_{lead}$,则 $Y_{lead} = Y_i$,此探狼代替头狼并发起召唤行为,否则此探狼退回原来的位置,继续向第二个方向前进,向第 p ($p = 1, 2$) 个方向前进进一步后所对应的位置为:

$$X_i^p = X_i + (-1)^p * stepa \quad (12)$$

此时探狼感知的猎物气味浓度为 Y_i^p ,选择其中一个对应函数值小于原始位置所对应函数值方向前进,更新该探狼的位置 X_i ,对游走行为迭代多次,直到探狼迭代到最大的游走次数或者某一只探狼对应的函数值 $Y_i < Y_{lead}$ 然后进入狼群算法的召唤行为。

(c) 召唤行为. 头狼向其它人工狼发起召唤行为,然后周围的 F_num ($F_num = Np - S_num - 1$) 匹猛狼会以相应的步长 $stepb$ 快速的向头狼所在的位置靠近,第 $k+1$ 次迭代后猛狼所处的位置为:

$$X_i^{k+1} = X_i^k + stepb * (g^k - X_i^k) \quad (13)$$

式中, g^k 为第 k 代群体中头狼所在的位置,在猛狼向头狼靠近的过程中,如果某个猛狼寻得的猎物所对应的函数值 $Y_i < Y_{lead}$,那么就对头狼进行更新. 此猛狼会成为新的头狼然后对其它猛狼发起召唤行为,否则此猛狼会继续向头狼靠近,直到该猛狼与头

狼 g^k 之间的 Manhattan^[14] 距离 dis 小于 $dnear$ 时,停止奔袭。

4.2 动态调节 DBN 结构算法步骤

本文主要根据狼群算法中的游走与召唤行为寻得使数据集中所有样本重构误差和最小时所对应的节点个数. 当重构误差较小说明当前层所对应的 RBM 结构较好,以此确定该层网络节点个数,从而确定当前 DBN 的结构,然后对当前 DBN 网络提取的特征进行 K-means 聚类研究,根据聚类研究结果判断是否需要增加 DBN 网络的层数,以得到更好的 DBN 网络结构,详细步骤如下所示:

步骤 1 初始化相关参数. 狼群总个数 N_p , 狼群个体 X_i ($i = 1, 2, 3, \dots, N_p$), X_i 为神经元的个数. 探狼的个数 S_num , DBN 结构每层最多神经元个数 $maxdim$ 与最少神经元个数 $mindim$, 游走步长 $stepa$, 召唤步长 $stepb$, 猛狼与头狼之间的最小距离 $dnear$, RBM 网络中学习率 ϵ .

步骤 2 读取数据集. 为了避免数据各维取值范围不一致对聚类结果影响^[15],对数据集 x 进行归一化操作,如式(14)所示:

$$x_{ij} = \frac{x_{ij}}{\max\{x_{ij}\}} \quad (i = 1, 2, \dots, n; k = 1, 2, \dots, m) \quad (14)$$

步骤 3 如式(15)所示,计算狼群个体所对应神经元组成网络的所有样本重构误差的平方和. 对狼群个体按照对应的 error 值从小到大进行排序,则 $g^k = X_1, g^k$ 为头狼.

$$error = \sum_{m=1}^M \sum_{n=1}^N (v_i - v_i')^2 \quad (15)$$

式中, M 表示数据集中样本的个数, N 表示样本的维度,表示数据集中第 i 个样本特征,表示样本重构的特征。

步骤 4 将头狼以外的最佳 S_num 匹人工狼看成是探狼,然后根据公式(12) 执行游走行为. 如果所有探狼达到最大寻找次数或者某只探狼 X_i 搜寻的猎物对应的函数值 Y_i 小于头狼所对应的函数值 Y_{lead} 然后进行步骤 5 操作。

步骤 5 人工猛狼根据公式(13) 向头狼迅速靠近,如果在向 头狼靠近的过程中某只人工猛狼搜寻的猎物对应的函数值 $Y_i < Y_{lead}$,那么该人工猛狼就代替前一代头狼成为新的头狼并且向其它猛狼发起召唤行为,否则该猛狼会继续向头狼靠近,直到此猛

狼与头狼的距离小于 d_{near} .

步骤 6 转步骤 4. 直到达到最大迭代次数. 得到的头狼所对应的神经元个数为狼群算法找到的当前层网络比较好的神经元个数.

步骤 7 根据步骤 6 得到的神经元数作为新增加一层网络的神经元个数, 从而确定 DBN 结构, 提取初始数据的主要特征, 运用 K-means 算法对主要特征数据进行聚类研究, 计算聚类准确率.

步骤 8 增加隐含层. 运用步骤 1~6 的方法确定增加的隐含层神经元的个数, 然后运用 K-means 算法进行聚类测试, 计算聚类准确率. 如果聚类准确率更好, 那么就转步骤 4, 否者停止增加隐含层数.

步骤 9 得到较好的 DBN 网络结构与较高的聚类研究结果.

5 实验结果与分析

5.1 数据集

Iris 数据集. 数据来源: UCI 数据库 (<http://archive.ics.uci.edu/ml/>).

5.2 聚类评价指标

准确率 (precision, P) 是指聚类正确的样本与该类原有样本数量之间的比值. 在信息检索中, 它是衡量算法性能优劣的一个重要指标, 因此本文选取准确率作为评价聚类性能的指标, 表达式如式 (16) 所示:

$$p = \sum_i m_i / \sum_i M_i \tag{16}$$

式中, M_i 指第 i 类中的样本数量, m_i 表示正确聚类到第 i 类中的样本数量. P 越大说明聚类效果越好也说明 DBN 结构较好.

5.3 相关参数

本实验硬件环境: PC 机 (CPU: 2.53GHz, 内存: 2Gb), 软件环境 windows 7 操作系统, 在 matlabR2014a 编译器上实现.

相关参数: $N_p = 10$, $S_num = 5$, $maxdim = 2$, $mindim = 1$, $stepa = 3$, $stepb = 2$, $d_{near} = 1$, $MaxT = 20$, $y_{zMax} = 2$, $\epsilon = 0.01$, $batchsize = 1$, $numepochs = 2000$, $momentum = 0.5$.

根据文献[16]可以看出, 如果 DBN 网络中隐含层神经元个数较少, 那么会导致隐含层神经元的激活概率较大, 这样通过网络提取原来数据的特征会出现同质化现象; 如果 DBN 网络中节点较少, 那

么 DBN 在训练时候时间复杂度较高, 同样不利于样本训练, 所以本文设定隐含层节点最少为 2 个, 最多为前一层网络节点两倍.

5.4 实验结果

根据 k-means 算法对 Iris 数据集进行聚类测试, k-means 算法运行 50 次, 计算出 P 的平均值为 88.53%.

Iris 数据集维数为 4, 所以设定输入节点个数为 4. 增加一个隐含层. 由于 DBN 模型是一种概率模型, 所以根据 3.2 中的方法对算法运行多次, 取其中对应适应度函数值最小时所对应的节点数作为当前隐含层的节点数. 根据 3.2 节中算法得出当隐含层神经元个数为 8 时, 重构误差较小. 则确定当前 DBM 网络结构, 如图 3 所示.

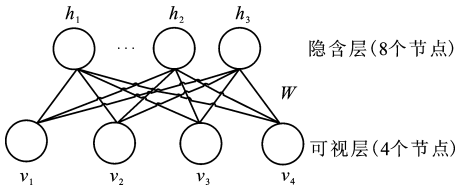


图 3 深度信念网络结构

当输入节点为 4, 输出节点为 8 时, 根据 DBN 模型提取 20 份原来数据的主要特征数据. 然后对每份数据进行 50 次 K-means 聚类测试, 计算平均聚类准确率. 对 20 份数据聚类测试得到的准确率如图 4 所示, 得到的准确率 P 的最大值是 95.09%, 平均值是 89.99%, 最小值是 86.67%.

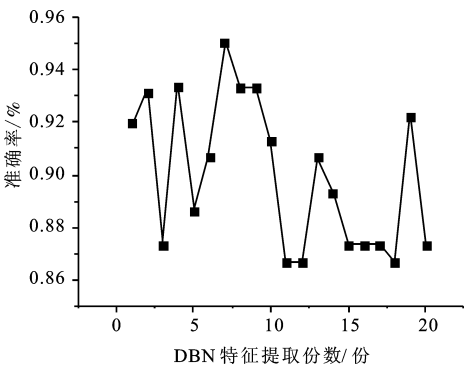


图 4 20 份特征数据对应的平均聚类准确率

增加一个隐含层, 根据 3.2 节中的方法对算法运行多次, 当神经元个数为 16 时, 重构误差较小, 则确定当前 DBN 结构如图 5 所示.

根据当前 DBN 网络结构, 提取 20 份原来数据的主要特征. 然后对每份特征数据进行 50 次 K-

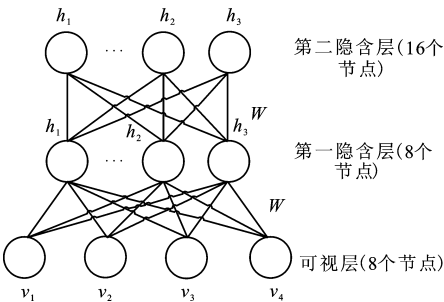


图 5 深度信念网络结构

means 聚类测试,计算出平均聚类准确率,如图 6 所示,得到的准确率 P 的最大值是 96.67%,平均值时 90.21%,最小值时 75.00%。

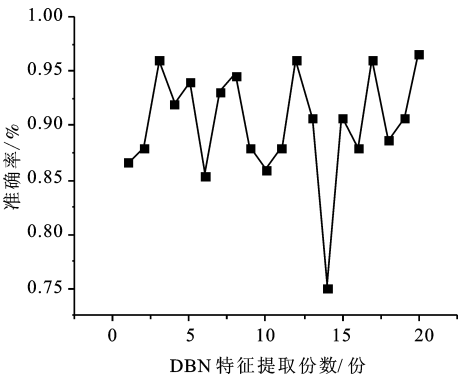


图 6 20 份特征数据对应的平均聚类准确率

再增加一个隐含层,根据 3.2 节方法对算法运行多次,当神经元个数为 21 时,重构误差较小,则确定当前 DBN 结构如图 7 所示。

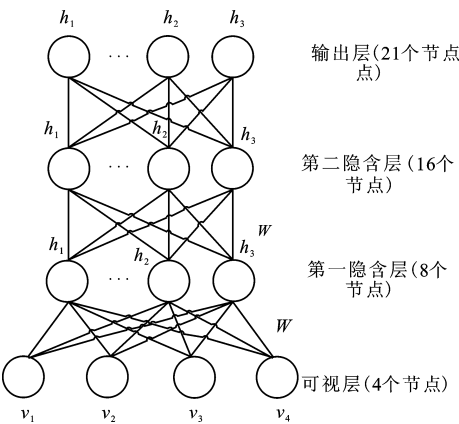


图 7 深度信念网络结构

根据当前 DBN 网络结构,提取 20 份原来数据的主要特征. 然后对每份特征数据进行 50 次 K-means 聚类测试,计算出平均聚类准确率. 准确率如图 8 所示. 得到的 P 的最大值是 98.00%,平均值时

80.15%,最小值时 66.67%。

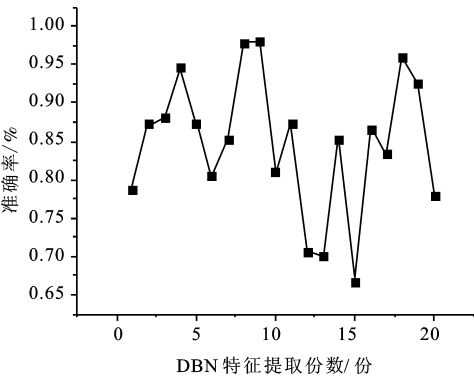


图 8 20 份特征数据对应的平均聚类准确率

当再增加一个隐含层,根据文本 3.2 节提出的 DBN 网络结构确定算法对算法运行多次,发现当神经元个数为 9 时,重构误差较小. 但是发现当增加的一层网络节点数为 9 时,所提取的特征对应的聚类准确率不是很好,得到的 P 的最大值为 93.01%。

以上可以看出,对于 Iris 数据集,当选择如图 7 所示的 DBN 网络结构模型时,能够提取更具有鉴别性的特征,并且聚类准确率最大为 98%,远大于对初始数据进行聚类测试计算得到的准确率 (88.53%)。

6 结束语

深度信念网络模型确定方法是一个开放性的研究方向,本文提出的基于狼群算法以最小化样本重构误差和为目标函数来确定 DBN 每层神经元的个数,在神经元个数确定的基础上,对 DBN 结构提取的原始数据的主要特征进行 K-means 聚类测试,根据聚类结果好坏来确定 DBN 结构的隐含层数是否增加,最终把聚类测试结果最好时的 DBN 结构作为所要确定的 DBN 结构. 根据此确定的 DBN 网络,输入数据从输入层输入,输出层可以得到更具鉴别性的特征,最终能够得到更好的聚类结果. 本文通过实验结果证明本文提出算法的可行性,但是当数据集比较大时,本文提出的 DBN 网络结构确定方法可能会面临时间复杂度比较大的困难,能否设计一个时间复杂度相对较低的算法来确定 DBN 网络结构将是接下来要研究的一个方向。

参考文献:

- [1] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786):504-507.
- [2] Bengio Y. Learning deep architectures for AI[J]. Foundations and trends in Machine Learning, 2009, 2(1):1-127.
- [3] Wang Xue. Intelligent testing of tests [M]. Beijing: Qinghua University Press, 2008.
- [4] Gao Daqi. Research on the structure of three layer neural network of teachers with linear basic function [J]. Journal of Circuits and Systems, 1997, 21(3): 31-37.
- [5] Benardos P G, Vosniakos G C. Optimizing feedforward artificial neural network architecture[M]. Oxford, Pergamon Press, Inc. 2007.
- [6] Hinton G E. A practical guide to training restricted boltzmann machines[J]. Momentum, 2012, 9(1):599-619.
- [7] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 1989, 18(7):1527-1554.
- [8] Schölkopf B, Platt J, Hofmann T. Greedy layer-wise training of deep networks[M]. Cambridge, Massachusetts, United States, MIT Press, 2007:153-160.
- [9] Hinton G E. Learning multiple layers of representation. [J]. Trends in Cognitive Sciences, 2007, 11(10):428-432.
- [10] Bengio Y, Delalleau O. Justifying and generalizing contrastive divergence[J]. Neural Computation, 1989, 21(6):1601-1605.
- [11] Li Fei, Guang GaoXiao, Wan Kaifang. Research on RBM training algorithm based on dynamic gibbs sampling [J]. Acta Automatica Sinica, 2016, 42(6): 931-942.
- [12] 吴虎胜, 张凤鸣, 吴庐山. 一种新的群体智能算法-狼群算法[J]. 系统工程与电子技术, 2013, 35(11): 2430-2438.
- [13] Salakhutdinov R, Murray I. On the quantitative analysis of deep belief networks[C]// International Conference on Machine Learning. ACM, 2008:872-879.
- [14] Tokdar S T, Kass R E. Importance sampling: a review[J]. Wiley Interdisciplinary Reviews Computational Statistics, 2010, 2(1):54-60.
- [15] Liu Liqing, Ding Qiaolin, Zhang Tiefeng. Influence of data preprocessing method on fuzzy C - means clustering [J]. Electric Power Science and Engineering, 2011, 27(8):24-27.
- [16] Liu Kai, Zhang Li-min, Sun Yongwei. RBM optimization design based on genetic algorithm [J]. Microelectronics and Computer, 2005, 35(21):660-662.

作者简介:

周贤泉 男, (1990-), 硕士研究生. 研究方向为数据挖掘、深度学习等. E-mail: 295387083@qq.com.

宋威 男, (1981-), 博士, 副教授. 研究方向为数据挖掘、人工智能、模式识别和信息检索.