

DOI: 10.11992/tis.201809018

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20190531.1405.003.html>

一种基于模糊划分和模糊加权的集成深度信念网络

张雄涛^{1,2}, 胡文军², 王士同¹

(1. 江南大学 数字媒体学院, 江苏 无锡 214122; 2. 湖州师范学院 信息工程学院, 浙江 湖州 313000)

摘要: 针对 DBN 算法训练时间复杂度高, 容易过拟合等问题, 受模糊理论启发, 提出了一种基于模糊划分和模糊加权的集成深度信念网络, 即 FE-DBN(ensemble deep belief network with fuzzy partition and fuzzy weighting), 用于处理大样本数据的分类问题。通过模糊聚类算法 FCM 将训练数据划分为多个子集, 在各个子集上并行训练不同结构的 DBN, 将每个分类器的结果进行模糊加权。在人工数据集、UCI 数据集上的实验结果表明, 提出的 FE-DBN 比 DBN 精度均有所提升, 具有更快的运行时间。

关键词: 集成; 深度信念网络; 模糊划分; 模糊加权; 运行时间; 模糊聚类算法; 模糊理论

中图分类号: O235; TP18 **文献标志码:** A **文章编号:** 1673-4785(2019)05-0905-10

中文引用格式: 张雄涛, 胡文军, 王士同. 一种基于模糊划分和模糊加权的集成深度信念网络 [J]. 智能系统学报, 2019, 14(5): 905-914.

英文引用格式: ZHANG Xiongtao, HU Wenjun, WANG Shitong. Ensemble deep belief network based on fuzzy partitioning and fuzzy weighting[J]. CAAI transactions on intelligent systems, 2019, 14(5): 905-914.

Ensemble deep belief network based on fuzzy partitioning and fuzzy weighting

ZHANG Xiongtao^{1,2}, HU Wenjun², WANG Shitong¹

(1. School of Digital Media, Jiangnan University, Wuxi 214122, China; 2. School of Information Engineering, Huzhou University, Huzhou 313000, China)

Abstract: Aiming at the problems of high training time complexity and easy over-fitting of the deep belief network (DBN) algorithm, inspired by the fuzzy theory, an ensemble deep belief network based on fuzzy partitioning and fuzzy weighting, namely FE-DBN (ensemble deep belief network with fuzzy partition and fuzzy weighting), is proposed to deal with the classification of large-scale data. First, the training data is divided into several subsets by fuzzy clustering algorithm (FCM), and then the DBNs of different structures are trained in parallel on each subset. Finally, the results of each classifier are ensemble by fuzzy weighting. Experiments on artificial datasets and UCI datasets show that the proposed FE-DBN outperforms the DBN in terms of accuracy and running time.

Keywords: ensemble; deep belief network; fuzzy partition; fuzzy weighting; running time; fuzzy clustering algorithm (FCM); fuzzy theory

近年来, 深度学习在图像识别和语音识别领域取得了突破性的进展。深度学习逐渐成为机器学习最热的研究方向之一。由于 RBM(restricted boltzmann machine)^[1-3] 具有表达能力强、易于推理等优点被成功用作深度神经网络的结构单元。当前, 以 RBM^[4-5] 为基本构成模块的 DBN(深度信念

网络)、DBM(深度玻尔兹曼机)等模型被认为是最有效的深度学习算法。其中深度信念网络(DBN)是深度学习的典型代表, 通常 DBN 在进行图像及语音等模式识别上有较高的精度, 但是训练一个 DBN 的复杂度非常高, 因为 DBN 在微调阶段使用了 BP 算法, 这种算法很难做到多机并行, 所以在大规模的数据上进行学习会非常困难。综上, DBN 主要存在两个问题: 1) 训练一个 DBN 的时间复杂度仍然较高; 2) 达到好的效果通

收稿日期: 2018-09-13. 网络出版日期: 2019-06-03.

基金项目: 国家自然科学基金面上项目 (61572236, 61300151, 61772198).

通信作者: 张雄涛. E-mail: 1047897965@qq.com.

常需要较多的隐含节点数,然而当隐含节点数较多时又易产生过拟合。虽然邓力等^[6-8]通过改进 DBN 的网络结构来提升其性能,但仍然没有突破以上问题。

分类是深度学习的核心问题,提高分类器的分类性能是分类器研究的主要目标之一。通常将模糊理论与分类器相结合,用于处理不确定问题。在构建分类模型时,为了对所构建的目标模型的局部细节有更好的逼近程度,通常的做法是将输入空间划分成多个模糊区域或模糊子空间,即模糊划分^[9-10],在各个子空间中分别构建分类子模型,最后将各子模型的结果集成输出^[11-12]。有 3 种模糊划分方法,即格状划分、树状划分和散状划分。格状划分是将每一维度的输入空间作划分,求得其模糊集合,再根据模糊系统理论,将模糊集映射成模糊区域。树状划分是,一次产生一个与模糊区域相对应的一个划分,每做一次划分就会产生出划分面。散状划分是,将输入输出的数据作分析,将预产生相似结果的输入空间以模糊区域作划分,每一模糊区域可作描述输入输出数据的行为。该划分是一种较为灵活的划分方法,吸收了前两种方法的优点,同时摒弃了它们存在的不足。在本文的研究中,采用散状划分这种方式。

为了更好地挖掘深度模型的表达能力,在实际应用中进一步提高 DBN 的精度并加快 DBN 的训练。受到上述思想的启发,本文提出了一种基于模糊划分和模糊加权的集成深度信念网络,将对应的集成分类算法命名为 FE-DBN。首先通过模糊聚类算法 FCM 将训练数据划分为多个子集;然后在各个子集上并行训练不同结构的 DBN;最后借鉴模糊集合理论的思想,将各个分类器的结果进行模糊加权。该算法能够有效且快速解决大样本数据的分类问题,克服了单个 DBN 用于数据分类时时间复杂度较高等缺点;而且,FE-DBN 可以避免过拟合问题,具有分类精度高等优点;并在人工数据集、UCI 数据集上得到了有效验证。

1 RBM 和 DBN

受限玻尔兹曼机是由 Hinton 和 Sejnowski 于 1986 年提出的一种生成式随机网络^[1],该网络是一种基于能量的概率图模型,它由一个可见层和一个隐含层组成,如图 1 所示, \mathbf{v} 和 \mathbf{h} 分别表示可见层与隐含层, \mathbf{W} 表示两层之间的连接权值。对于可见层与隐含层,其连接关系为层间全连接,层内无连接。注意,图 1 中 \mathbf{h} 有 m 个节点, \mathbf{v} 有 n 个节点,单个节点用 v_i 和 h_j 描述。可见层用于观测数据,隐含层用于提取特征。RBM 的隐单元

和可见单元可以为任意的指数族单元。本文只讨论所有的可见层和隐含层单元均为伯努利分布,假设所有的可见单元与隐单元均为二值变量,即对 $\forall i, j, v_i \in \{0, 1\}, h_j \in \{0, 1\}$ 。

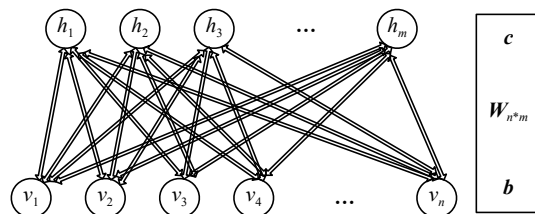


图 1 RBM 示意图

Fig. 1 The structure of RBM

RBM 是一种能量模型,能量函数定义为

$$E(\mathbf{v}, \mathbf{h} | \boldsymbol{\theta}) = - \sum_{i=1}^n b_i v_i - \sum_{j=1}^m c_j h_j - \sum_{i=1}^n \sum_{j=1}^m v_i W_{ij} h_j$$

式中: \mathbf{b} 和 \mathbf{c} 分别为可见层和隐含层的偏置向量; $\boldsymbol{\theta} \in \{\mathbf{b}, \mathbf{c}, \mathbf{W}\}$, \mathbf{W} 表示权值矩阵。基于能量函数,可以得到 \mathbf{v} 和 \mathbf{h} 的联合概率分布:

$$\begin{cases} P(\mathbf{v}, \mathbf{h}, \boldsymbol{\theta}) = \frac{e^{-E(\mathbf{v}, \mathbf{h}, \boldsymbol{\theta})}}{Z} \\ Z = \sum_{\tilde{\mathbf{v}}} \sum_{\tilde{\mathbf{h}}} e^{-E(\tilde{\mathbf{v}}, \tilde{\mathbf{h}}, \boldsymbol{\theta})} \end{cases} \quad (1)$$

式中 Z 函数为归一项。

多个 RBM 的堆栈组合构成了 DBN^[5], 前一个 RBM 的输出作为后一个 RBM 的输入。如图 2 所示,最底层是输入层,最顶层是输出层,中间层是隐含层。DBN 的学习包括两个阶段:预训练和微调。预训练是以贪婪的无监督的方式逐层进行训练的,将输入层映射到输出层从而学习到复杂的非线性函数;微调是在监督的方式下实现的,它使用反向传播 (BP) 算法从最顶层到最底层对整个 DBN 网络参数进行微调。

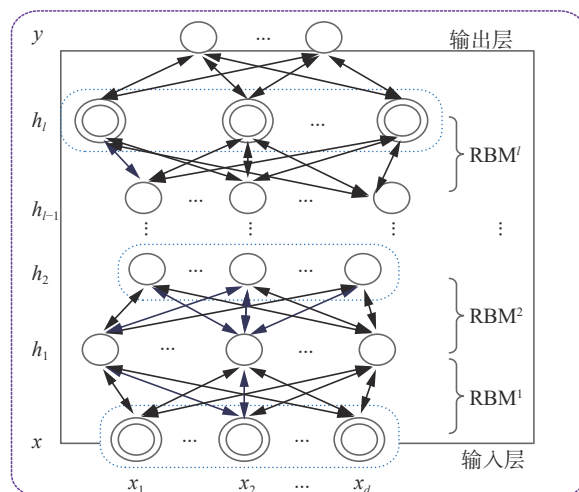


图 2 DBN 结构

Fig. 2 The structure of DBN

尽管 DBN 具有强大的知识表达能力, 但是当处理大规模数据甚至大数据时, DBN 在微调阶段需要花费大量的时间去训练模型, 这样导致训练时间特别长。

2 基于模糊划分和模糊加权的 DBN 分类器集成

针对不同的子空间, 从不同的角度来刻画一个物体各个方面的性质, 利用各种不同结构分类器的组合来形成一个综合的决策。大量的实验和应用证明: 将多个分类器的决策结果按照一定的规则集成在一起, 往往可以得到比其中最优分类器还要好的性能^[12-15]。针对经典的 DBN 算法时间复杂度高, 容易过拟合, 本文将具有多个不同隐含层结构的 DBN 分类器进行集成, 提出了一种基于模糊划分和模糊加权的集成 DBN, 即 FE-DBN。提出的 FE-DBN 性能优于经典的 DBN 分类算法, 主要有两个原因: 1) 在分类之前首先进行预处理、聚类, 能够更好地增强分类精度; 2) 将训练集分为多个子集, 每个子集用不同结构的 DBN 训练, 最后将结果进行模糊加权。根据集成学习原理, 多个弱分类器的组合能够组成一个强分类器。

2.1 FE-DBN 结构图

FE-DBN 结构图如图 3 所示, 首先利用模糊聚类算法 FCM, 将训练数据集划分为 K 个子集, 每个子集分别采用不同结构的 DBN 模型进行建模 (每个 DBN 子模型中每层隐节点数不一样, 由此构成了 K 个 DBN 模型), 各模型独立并行训练, 最后将各模型所得结果进行模糊加权形成最终输出。在进行模糊加权时, 采用高斯型隶属度函数进行权值计算。在 FE-DBN 中, 各 DBN 子模型并行训练, 由于各个训练子集的数据规模远远小于原数据规模, 需要较少的隐节点数, 因此训练时间较短。

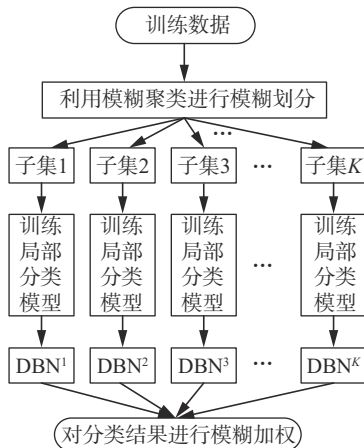


图 3 FE-DBN 结构

Fig. 3 The structure of FE-DBN

2.2 实现过程

首先, 使用模糊聚类算法 FCM 对训练数据集进行模糊分组。利用 FCM 算法进行模糊聚类, FCM 的目标函数为^[16-19]:

$$\min_{\mu, v} J = \sum_{i=1}^K \sum_{j=1}^N \mu_{ij}^m (\|x_j - v_i\|^2)$$

$$\text{s.t. } \sum_{i=1}^K \mu_{ij}^m = 1, \forall j = 1, 2, \dots, n$$

式中: K 是划分数; N 是样本数; $v_i = (v_{i1}, v_{i2}, \dots, v_{id})$ 为第 i 类的中心点; μ_{ij} 表示第 j 个样本属于第 i 类的隶属度; m 是模糊指数, 必须满足 $m \geq 2$, 本文取为 2; x_j 表示第 j 个样本点。引入拉格朗日因子构造新的目标函数, 推导得到隶属度和聚类中心的迭代计算公式分别为^[20]:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^K \left(\frac{\|x_j - v_i\|^2}{\|x_j - v_k\|^2} \right)^{1/(m-1)}} \quad (2)$$

$$v_i = \sum_{j=1}^N \mu_{ij}^m x_j / \sum_{j=1}^N \mu_{ij}^m, 1 \leq i \leq K \quad (3)$$

根据式 (2)、式 (3), 当迭代终止后, 所获得的隶属矩阵 U 在去模糊化后便得到空间划分矩阵。

根据式 (2)、式 (3), 计算宽度为

$$\gamma_j^s = \sqrt{\frac{\sum_{i=1}^N \mu_{ij}^m \cdot \|x_i^s - v_j^s\|^2}{\sum_{i=1}^N \mu_{ij}^m}}$$

根据聚类中心和宽度的值, 并利用式 (4) 对训练数据集进行模糊划分:

$$\vartheta_j = \{(x_i, y_i) | v_j^s - \xi \cdot \gamma_j^s \leq x_i^s \leq v_j^s + \xi \cdot \gamma_j^s\}$$

$$s = 1, 2, \dots, q; j = 1, 2, \dots, K \quad (4)$$

式中: ϑ_j 为定义的划分子集; q 为维数; ξ 为重叠因子, ξ 越大子集划分也就越模糊。模糊划分示意图如图 4 所示。在对原数据集模糊划分完成后, 再在各个子集上并行训练不同结构的 DBN。

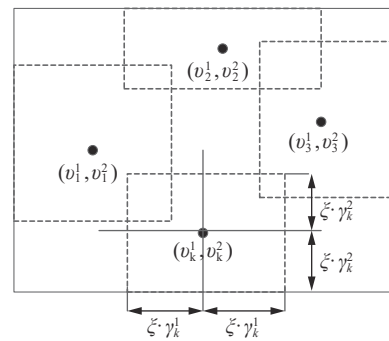


图 4 模糊划分示意

Fig. 4 Fuzzy partition

在式(1)中最关心的是联合概率分布所确定的边缘概率分布 $P(\mathbf{v}|\boldsymbol{\theta})$, 由于 RBM 模型层内无连接, 因此当给定可见单元的状态时, 各隐单元的激活状态是条件独立的。此时, 第 j 个隐单元的激活概率为

$$P(h_j = 1|\mathbf{v}, \boldsymbol{\theta}) = \sigma(c_j + \sum_{i=1}^n v_i W_{ij}) \quad (5)$$

式中 $\sigma(x) = \frac{1}{1 + \exp(-x)}$ 为 sigmoid 激活函数。第 i 个可见单元的激活概率为

$$P(v_i = 1|\mathbf{h}, \boldsymbol{\theta}) = \sigma(b_i + \sum_{j=1}^m W_{ij} h_j) \quad (6)$$

RBM 采用 Hinton 提出的 CD- k (对比散度) 算法进行参数学习, 并证明, 当使用训练样本初始化 $\mathbf{v}^{(0)}$ 时, 仅需较少的抽样步数(一般 $k=1$) 就可以得到很好的近似。采用 CD- k 算法, 各参数的更新准则如下^[3]:

$$\begin{aligned} \Delta W_{ij} &= \varepsilon(<v_i h_j>_{\text{data}} - <v_i h_j>_{\text{recon}}) \\ \Delta b_i &= \varepsilon(<v_i>_{\text{data}} - <v_i>_{\text{recon}}) \\ \Delta c_j &= \varepsilon(<h_j>_{\text{data}} - <h_j>_{\text{recon}}) \end{aligned} \quad (7)$$

式中: ε 为预训练的学习率; $<\cdot>_{\text{data}}$ 为训练数据集所定义的分布之上的数学期望; $<\cdot>_{\text{recon}}$ 为重构后的模型所定义的分布上的期望。利用式(7), 通过迭代更新得到 DBN 各子模型的参数。各子模型训练好之后, 给定测试数据 \mathbf{x}_i , 计算得到该数据在每个模型的输出结果, 利用三角形隶属度函数计算权值:

$$\omega_k(\mathbf{x}_i) = \omega_k^1(\mathbf{x}_i^1) \omega_k^2(\mathbf{x}_i^2) \cdots \omega_k^q(\mathbf{x}_i^q) \quad (8)$$

$$\omega_k^s(\mathbf{x}_i^s) = \max \left\{ \min \left(\frac{x_i^s - (v_k^s - \xi \cdot \gamma_k^s)}{v_k^s - (v_k^s - \xi \cdot \gamma_k^s)}, \frac{(v_k^s + \xi \cdot \gamma_k^s) - x_i^s}{(v_k^s + \xi \cdot \gamma_k^s) - v_k^s} \right), 0 \right\} \quad (9)$$

式中: $k = 1, 2, \dots, K; i = 1, 2, \dots, N; s = 1, 2, \dots, q$ 。

划分好样本空间, 每一个分类器在样本子空间进行运算, 样本在分类器中具有局部分类性能最好的, 其所对应的权值就越大。

最后, 将各 DBN 分类器所得结果进行模糊加权, 即

$$\hat{y}(\mathbf{x}_i) = \frac{\sum_{k=1}^K \omega_k(\mathbf{x}_i) \text{LCM}_{\text{DBN}}^k(\mathbf{x}_i)}{\sum_{k=1}^K \omega_k(\mathbf{x}_i)} \quad (10)$$

式中: $\text{LCM}_{\text{DBN}}^k(\mathbf{x}_i)$ 为样本 \mathbf{x}_i 在第 k 个模型的分类结果; LCM 为局部分类模型; \hat{y} 为 K 个模型分类

结果模糊加权后所得的最终输出。

FE-DBN 算法实现过程如下:

1) 初始化。设定划分子集个数 K 及重叠因子 ξ , 各子模型 DBN 的隐节点数及 DBN 的迭代周期, 初始化 $\mathbf{W}, \mathbf{b}, \mathbf{c}$ 的值, 学习率 ε 。

2) 划分子集。利用模糊聚类算法 FCM 求得每簇的中心点和宽度, 根据式(4)将源数据集划分为 K 个子集。

3) 并行训练各子模型 $\text{DBN}^1 \sim \text{DBN}^K$, 对于所有的可见单元, 利用式(5)计算 $P(h_j = 1|\mathbf{v}, \boldsymbol{\theta})$, 并抽取 h_j 且 $h_j \in \{0, 1\}$ 对于所有的隐单元, 利用式(6)计算 $P(v_i = 1|\mathbf{h}, \boldsymbol{\theta})$, 并抽取 v_i 且 $v_i \in \{0, 1\}$, 利用式(7)更新 RBM 参数 $\mathbf{W}, \mathbf{b}, \mathbf{c}$ 的值, 即

$$\mathbf{W} = \mathbf{W} + \Delta \mathbf{W}, \mathbf{b} = \mathbf{b} + \Delta \mathbf{b}, \mathbf{c} = \mathbf{c} + \Delta \mathbf{c}$$

重复 3), 直到满足迭代周期为止。

4) 利用式(8)、式(9)计算每个测试数据对各个子集的隶属度, 将测试数据代入 3) 所得的 K 个子模型中并输出 K 个分类结果。利用式(10)进行集成得到最终输出。

3 实验与分析

本文在实验部分将分别利用人工数据和 UCI 数据对所提的基于模糊划分和模糊加权的集成 DBN 分类算法 (FE-DBN) 进行验证和评估。并将该算法的性能同深度信念网络 (DBN)^[1-3] 算法进行比较。为了验证本文所提出的算法 FE-DBN 的有效性, 采用的对比算法有局部分类模型 DBN^K 和全局分类模型 DBN, 其中 DBN^K 表示将原数据集分为 K 个子集, 在每个子集上构建一个局部深度信念网络分类模型。所有的实验结果都采用五折交叉, 运行 10 次取均值。

3.1 实验设置

3.1.1 数据集

人工数据集生成两种: 左螺旋型、右高斯型, 如图 5。两种数据均生成 4 000 个样本, 螺旋型 2 类, 2 维; 高斯型 4 类, 2 维。构造的螺旋型数据集正负类样本数各 2 000, 高斯型数据集每类样本数 1 000, 高斯型各类的中心分别是: [7 8]、[15 13]、[15 5]、[23 8], 协方差均为 $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ 。真实数据集全部来自于 UCI^[21]。数据集详细信息如表 1、表 2 所示。

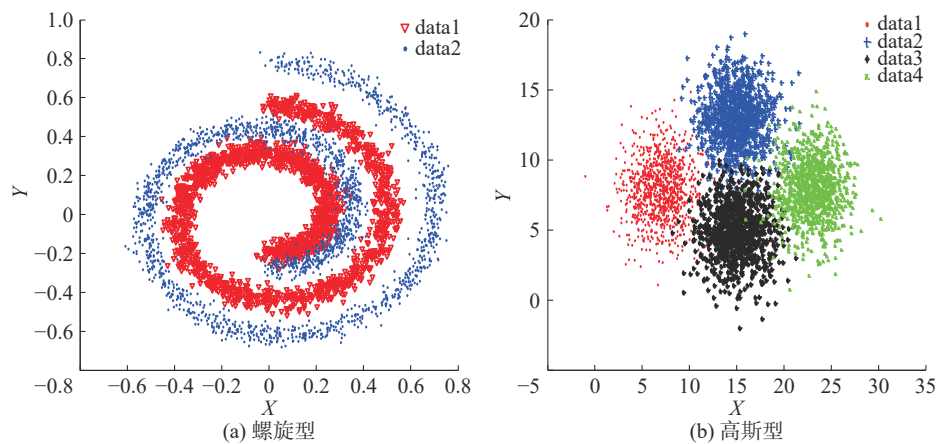


图5 人工数据集

Fig. 5 Artificial datasets

表1 人工数据集
Table 1 Artificial datasets

数据集	类别数	特征数	样本数
Swiss	2	2	4 000
Gauss	4	2	4 000

表2 UCI 数据集
Table 2 UCI datasets

数据集	类别数	特征数	样本数
Adult	2	15	48 841
Magic_gamma_telescope	2	11	19 020
pendigits	10	16	7 494
Waveform3	3	21	5 000
shuttle	7	8	58 000

3.1.2 参数设置及实验运行环境

实验中采用三层的 DBN, ξ 用于控制子集的伸缩宽度, 在实验中发现, 当 $\xi=3$ 时, 能够取得较好的结果, 也可根据具体的数据集分布进行微

调。DBN 代码参照 <http://www.cs.toronto.edu/~hinton/>, RBM 迭代周期 maxepoch=20, 用于控制 RBM 的预训练迭代次数和模型参数的微调次数。权重的学习率 epsilonw=0.05; 显层偏置的学习率 epsilonvb=0.05; 隐层偏置的学习率 epsilonhb=0.05; 权损失系数 weightcost=0.000 2; 动量学习率 initialmomentum=0.5, finalmomentum=0.9。

本文使用平均测试精度、均方差、运行时间(训练时间+测试时间)进行算法性能度量。实验环境为 intel(R) Core(TM) i3 3.40 GHz CPU, 8 GB 内存, Windows10 操作系统, MATLAB2016a。

3.2 实验结果及分析

为进一步探索数据集模糊划分个数对提升分类精度及算法运行时间的重要性, 本文将数据集划分为不同的子集个数, 及采用不同的隐节点数组合分别进行实验比较。如表 3 所示, 局部分类模型 DBN^K 分别有 3 个子集和 4 个子集, “28+22+19”表示 DBN¹ 中第一层、第二层、第三层的隐节点数分别为 28、22、19。

表3 在 Swiss 数据集上的分类精度及运行时间对比

Table 3 Performance in terms of average testing accuracy and running time with their standard deviation on Swiss

算法	子集K	隐节点个数	测试精度	运行时间/s
局部分类模型DBN ^K (k=3)	1	28+22+19	0.515 0±0.021 5	3.223 9±0.269 2
	2	30+27+24	0.521 3±0.033 0	3.941 8±0.165 3
	3	38+32+26	0.517 5±0.035 9	4.796 0±0.232 0
FE-DBN (k=3)	—	—	0.535 8±0.017 5	5.304 1±0.327 5
DBN (m=4)	—	40+40+40	0.535 6±0.015 0	6.609 8±0.310 6
局部分类模型DBN ^K (k=4)	1	30+20+10	0.520 6±0.029 7	2.510 5±0.358 2
	2	23+23+14	0.519 7±0.036 6	2.612 7±0.211 3
	3	25+19+16	0.522 2±0.039 2	2.611 8±0.255 1
	4	23+20+17	0.524 5±0.035 8	2.779 1±0.283 6
FE-DBN (k=4)	—	—	0.545 0±0.026 3	3.973 8±0.227 5
DBN (m=30)	—	30+30+30	0.535 3±0.029 1	4.354 2±0.230 0

3.2.1 人工数据集

该实验部分主要是通过构造模拟数据集来验证本文提出的 FE-DBN 算法的有效性。从表 3、表 4 的实验结果可以看出:螺旋线数据集不太好

区分,精度不高,但是 FE-DBN 仍有所提升;高斯型数据集精度,FE-DBN 比各局部模型 DBN^k 略高,和全局模型 DBN 基本持平,因为其精度已经很高,故很难再有较大的提升。

表 4 在 Gauss 数据集上的分类精度及运行时间对比

Table 4 Performance in terms of average testing accuracy and running time with their standard deviation on Gauss

算法	子集 K	隐节点个数	测试精度	运行时间/s
局部分类模型 $\text{DBN}^k (k=3)$	1	28+23+18	0.929 3±0.004 4	2.823 6±0.122 2
	2	31+25+20	0.923 8±0.004 8	2.518 9±0.182 9
	3	25+21+17	0.935 0±0.029 1	2.241 8±0.157 5
FE-DBN ($k=3$)	—	—	0.942 5±0.020 3	3.185 4±0.263 4
DBN ($m=35$)	—	35+35+35	0.940 6±0.003 1	4.116 7±0.357 8
局部分类模型 $\text{DBN}^k (k=4)$	1	20+15+10	0.921 2±0.001 7	1.735 8±0.167 3
	2	22+16+10	0.921 6±0.001 5	1.719 8±0.121 3
	3	24+17+10	0.919 0±0.002 2	1.744 8±0.192 0
	4	28+23+18	0.937 1±0.011 5	2.342 4±0.192 3
FE-DBN ($k=4$)	—	—	0.948 8±0.009 1	2.815 4±0.166 1
DBN ($m=30$)	—	30+30+30	0.949 5±0.008 3	3.298 6±0.121 8

3.2.2 UCI 数据集

本部分实验选的 UCI 数据集,既有中等规模

数据,又有大规模数据,既有二分类,也有多分类,3 种算法在各 UCI 数据集上的对比实验结果如表 5~9 所示。

表 5 在 Adult 数据集上的分类精度及运行时间对比

Table 5 Performance in terms of average testing accuracy and running time with their standard deviation on Adult

算法	子集 K	隐节点个数	测试精度	运行时间/s
局部分类模型 $\text{DBN}^k (k=3)$	1	70+60+50	0.826 6±0.044 6	80.368 0±3.75
	2	70+60+60	0.828 4±0.011 7	90.170 0±3.45
	3	75+65+55	0.823 1±0.048 9	89.318 2±7.65
FE-DBN ($k=3$)	—	—	0.835 7±0.028 4	95.901 2±7.69
DBN ($m=70$)	—	70+70+70	0.829 9±0.019 3	105.050 5±3.32
局部分类模型 $\text{DBN}^k (k=4)$	1	65+60+50	0.820 8±0.014 7	80.501 8±5.17
	2	60+60+50	0.822 2±0.024 3	78.606 8±5.76
	3	60+55+45	0.819 7±0.022 5	72.549 1±4.99
	4	60+55+50	0.820 7±0.027 3	81.445 7±4.43
FE-DBN ($k=4$)	—	—	0.841 7±0.014 1	83.607 6±3.72
DBN ($m=60$)	—	60+60+60	0.830 6±0.001 2	88.826 9±2.75
局部分类模型 $\text{DBN}^k (k=5)$	1	40+30+20	0.820 6±0.012 1	36.562 9±2.67
	2	45+35+25	0.817 0±0.014 5	41.957 9±3.01
	3	50+40+30	0.815 0±0.023 5	51.769 7±5.23
	4	40+30+30	0.817 7±0.002 3	46.104 1±2.40
	5	40+40+40	0.818 7±0.002 0	55.010 6±1.66
FE-DBN ($k=5$)	—	—	0.848 6±0.014 1	59.097 9±3.28
DBN ($m=50$)	—	50+50+50	0.829 5±0.006 3	73.346 4±2.88

表 6 在 Magic_gamma_telescope 数据集上的分类精度及运行时间对比

Table 6 Performance in terms of average testing accuracy and running time with their standard deviation on Magic_gamma_telescope

算法	子集 K	隐节点个数	测试精度	运行时间/s
局部分类模型 $\text{DBN}^K(k=3)$	1	55+45+45	0.788 5±0.001 5	28.341 1±1.66
	2	50+50+40	0.789 3±0.002 8	28.732 1±1.35
	3	50+45+40	0.787 9±0.002 0	26.759 7±2.37
FE-DBN ($k=3$)	—	—	0.801 3±0.013 9	31.598 4±2.28
DBN ($m=60$)	—	60+60+60	0.791 0±0.019 5	35.997 3±2.08
局部分类模型 $\text{DBN}^K(k=4)$	1	40+30+30	0.787 5±0.001 5	22.240 0±0.92
	2	40+30+20	0.787 1±0.001 9	17.611 2±1.18
	3	45+35+25	0.788 4±0.008 6	20.656 6±2.34
	4	40+40+40	0.788 3±0.008 2	26.091 8±2.52
FE-DBN ($k=4$)	—	—	0.807 1±0.009 4	28.568 9±1.70
DBN ($m=50$)	—	50+50+50	0.789 8±0.011 8	30.589 6±0.28
局部分类模型 $\text{DBN}^K(k=5)$	1	30+30+20	0.782 5±0.005 3	18.708 9±0.89
	2	30+25+25	0.785 3±0.001 2	18.242 3±0.74
	3	35+30+30	0.785 6±0.002 5	21.759 0±0.65
	4	30+25+20	0.782 2±0.006 2	17.100 7±0.53
	5	30+30+30	0.781 9±0.001 2	21.417 0±0.76
FE-DBN ($k=5$)	—	—	0.811 6±0.003 9	23.090 8±0.61
DBN ($m=40$)	—	40+40+40	0.789 6±0.006 2	25.952 5±0.35

表 7 在 pendigits 数据集上的分类精度及运行时间对比

Table 7 Performance in terms of average testing accuracy and running time with their standard deviation on pendigits

算法	子集 K	隐节点个数	测试精度	运行时间/s
局部分类模型 $\text{DBN}^K(k=3)$	1	65+65+55	0.745 0±0.024 8	12.121 2±0.52
	2	60+55+50	0.745 8±0.003 4	10.503 4±0.79
	3	60+60+50	0.749 8±0.002 9	9.708 7±0.30
FE-DBN ($k=3$)	—	—	0.759 2±0.010 3	11.562 9±0.43
DBN ($m=65$)	—	65+65+65	0.758 5±0.007 8	12.431 3±0.85
局部分类模型 $\text{DBN}^K(k=4)$	1	50+50+45	0.735 7±0.030 2	8.833 2±0.29
	2	55+55+45	0.723 8±0.053 4	9.214 7±0.86
	3	50+45+45	0.713 9±0.043 4	8.874 0±0.89
	4	55+50+45	0.727 0±0.036 6	8.954 1±0.48
FE-DBN ($k=4$)	—	—	0.761 0±0.032 8	9.648 4±0.35
DBN ($m=55$)	—	55+55+55	0.756 5±0.028 4	10.384 6±0.77

表 8 在 Waveform3 数据集上的分类精度及运行时间

Table 8 Performance in terms of average testing accuracy and running time with their standard deviation on Waveform3

算法	子集 K	隐节点个数	测试精度	运行时间/s
局部分类模型 $\text{DBN}^K (k=3)$	1	60+60+50	0.863 0±0.014 9	8.851 6±0.38
	2	60+50+45	0.864 4±0.009 8	7.932 6±0.88
	3	60+55+45	0.863 2±0.029 4	7.724 1±0.54
FA_RBF ($k=3$)	—	—	0.871 0±0.016 0	9.069 3±0.25
DBN ($m=60$)	—	60+60+60	0.867 3±0.031 3	9.575 4±0.30
局部分类模型 $\text{DBN}^K (k=4)$	1	45+45+40	0.862 8±0.006 1	6.745 9±0.77
	2	50+50+40	0.863 2±0.015 3	7.275 0±0.61
	3	45+40+40	0.860 8±0.007 3	6.598 6±0.25
	4	50+45+40	0.863 0±0.015 1	6.836 9±0.40
FE-DBN ($k=4$)	—	—	0.874 5±0.030 2	7.547 5±0.49
DBN ($m=50$)	—	50+50+50	0.867 1±0.003 9	7.925 7±0.33

表 9 在 shuttle 数据集上的分类精度及运行时间

Table 9 Performance in terms of average testing accuracy and running time with their standard deviation on shuttle

算法	子集 K	隐节点个数	测试精度	运行时间/s
局部分类模型 $\text{DBN}^K (k=3)$	1	100+90+80	0.786 0±0.006 4	103.294 9±4.73
	2	100+90+90	0.798 3±0.030 1	119.187 7±8.71
	3	95+90+85	0.783 6±0.002 9	109.018 7±3.66
FE-DBN ($k=3$)	—	—	0.799 3±0.001 4	122.594 3±4.81
DBN ($m=100$)	—	100+100+100	0.799 5±0.005 7	130.596 0±3.31
局部分类模型 $\text{DBN}^K (k=4)$	1	85+75+65	0.778 4±0.002 4	90.773 3±3.07
	2	80+70+70	0.780 2±0.002 4	92.571 3±4.31
	3	95+85+75	0.788 5±0.001 2	102.628 3±2.52
	4	80+80+80	0.782 5±0.002 4	97.701 5±3.27
FE-DBN ($k=4$)	—	—	0.814 6±0.006 2	112.058 2±2.72
DBN ($m=90$)	—	90+90+90	0.794 3±0.044 6	120.630 9±8.54
局部分类模型 $\text{DBN}^K (k=5)$	1	70+60+50	0.780 5±0.002 6	64.656 8±3.19
	2	75+65+55	0.782 2±0.003 1	71.017 0±2.27
	3	80+70+60	0.781 3±0.001 3	83.305 8±4.51
	4	60+60+60	0.779 6±0.034 3	68.813 3±2.68
	5	65+65+65	0.773 5±0.012 9	73.431 0±3.08
FE-DBN ($k=5$)	—	—	0.806 7±0.006 9	92.183 0±1.95
DBN ($m=80$)	—	80+80+80	0.783 0±0.015 7	103.289 5±2.89

从表 5~9 的实验结果, 可以得出如下结论:

1) 在测试精度上, 和全局分类模型 DBN 相比, FE-DBN 在数据集 Adult、shuttle 和 Magic_gamma_telescope 上增长较多, 在数据集 pendigits 和 waveform3 上有略微上浮。在样本划分子

集确定的情况下, FE-DBN 高于任何一个局部分类模型 DBN^K 。总的来看, FE-DBN 算法的分类效果在三者中为最优。从表 5~9 中还可以看出, 在划分的子集数确定时, 具有不同隐节点数组合的各局部分类模型 DBN^K 分类器的精度并无明显差

异。随着划分子集数的增加, FE-DBN 的精度在不同数据集上基本均有增长的趋势。其主要原因在于, 根据集成原理^[14,22-23], 对于集成 FE-DBN 分类模型, 增加各子模型的多样性, 能够提高集成分类器的性能。

2) 和全局模型 DBN 相比较, FE-DBN 中每个局部分类模型需要较少的隐节点数, 就可以达到较高的精度, 这主要是因为组成 FE-DBN 的每个局部分类器都是弱分类器。

3) 对于所有数据集, 在运行时间上, 当划分子集数逐渐增多时, 由于每个子集的样本数在减少, 隐节点数也在减少, 运行时间相应也会减少。由于要进行模糊划分和模糊集成, FE-DBN 的运行时间比各局部分类模型 DBN^K 要多, 但是 FE-DBN 的运行时间要小于全局模型 DBN 的运行时间, 因为在 FE-DBN 中各局部分类模型是并行运行的, 且每个子模型的隐节点数均小于全局模型 DBN 的隐节点数。

无论是模拟数据集还是 UCI 数据集, 基于模糊划分和模糊加权的 DBN 集成分类器 (FE-DBN) 比单分类器 (DBN) 的性能好, 比最优的局部分类模型 DBN^K 也要高。由表 3~9 中结果, 根据统计分析得出, 样本划分粒度越细, 分类精度会越高, 表明细划分能得到更多的样本特征信息。但也不是子集划分得越多, 精度就越高, 数据集 shuttle 在划分子集数为 4 时取得最大值。

4 结束语

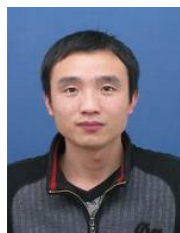
采用集成的方法解决 DBN 训练时间复杂度高的问题。根据数据之间的相似性信息对数据进行模糊分组, 构造样本空间子集, 然后在各样本空间子集中训练具有不同结构的 DBN 子分类器, 最后使用模糊加权的方法, 得到最终的集成分类器和分类结果。人工数据集和 UCI 数据集上的实验结果显示, FE-DBN 算法可以得到比其他分类算法更好的分类结果。未来的研究工作将主要集中在如何确定样本空间子集上。

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] HINTON G E. A practical guide to training restricted boltzmann machines[M]//MONTAVON G, ORR G B, MÜLLER K R. Neural Networks: Tricks of the Trade. Berlin, Heidelberg: Springer, 2012: 599-619.
- [3] HINTON G E, OSINDER S, TEH Y W. A fast learning algorithm for deep belief nets[J]. *Neural computation*, 2006, 18(7): 1527-1544.
- [4] DESJARDINS G, COURVILLE A, BENGIO Y. Adaptive parallel tempering for stochastic maximum likelihood learning of RBMs[C]//Proceedings of NIPS 2010 Workshop on Deep Learning and Unsupervised Feature Learning. Granada, Spain, 2010.
- [5] LEE S, CHANG J H. Deep belief networks ensemble for blood pressure estimation[J]. *IEEE access*, 2017, 5: 9962-9972.
- [6] QIU Xuecheng, ZHANG Le, REN Ye, et al. Ensemble deep learning for regression and time series forecasting[C]//Proceedings of 2014 IEEE Symposium on Computational Intelligence in Ensemble Learning. Orlando, USA, 2014: 1-6.
- [7] ZHANG Chong, LIM P, QIN A K, et al. Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics[J]. *IEEE transactions on neural networks and learning systems*, 2017, 28(10): 2306-2318.
- [8] MOHAMED A R, YU Dong, DENG Li. Investigation of full-sequence training of deep belief networks for speech recognition[C]//Proceedings of Interspeech 2010. Chiba, Japan, 2010: 2846-2849.
- [9] AU W H, CHAN K C C, WONG A K C. A fuzzy approach to partitioning continuous attributes for classification[J]. *IEEE transactions on knowledge and data engineering*, 2006, 18(5): 715-719.
- [10] CHUANG C C. Fuzzy weighted support vector regression with a fuzzy partition[J]. *IEEE transactions on systems, man, and cybernetics, part B*, 2007, 37(3): 630-640.
- [11] RATHORE P, BEZDEK J C, ERFANI S M, et al. Ensemble fuzzy clustering using cumulative aggregation on random projections[J]. *IEEE transactions on fuzzy systems*, 2018, 26(3): 1510-1524.
- [12] YU Zhiwen, ZHANG Yidong, CHEN C L P, et al. Multiobjective semisupervised classifier ensemble[J]. *IEEE transactions on cybernetics*, 2019, 49(6): 2280-2293.
- [13] ZHOU Ta, ZHANG Zhifeng, WANG Sifan, et al. Stacked unitwise fast combination classification for single hidden layer neural networks using extreme learning machine on EEG signals[J]. *Journal of medical imaging and health informatics*, 2018, 8(7): 1491-1495.
- [14] KUNCHEVA L I, WHITAKER C J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy[J]. *Machine learning*, 2003, 51(2): 181-207.
- [15] KITTLER J, HATEF M, DUIN R P W, et al. On combining classifiers[J]. *IEEE transactions on pattern analysis*

- and machine intelligence, 1998, 20(3): 226–239.
- [16] 蒋亦樟, 朱丽, 刘丽, 等. 多视角模糊双加权可能性聚类算法[J]. 智能系统学报, 2017, 12(6): 806–815.
- JIANG Yizhang, ZHU Li, LIU Li, et al. Multi-view fuzzy double-weighting possibility clustering algorithm[J]. CAAI transactions on intelligent systems, 2017, 12(6): 806–815.
- [17] LIU Jiefang, CHUNG F L, WANG Shitong. Black hole entropic fuzzy clustering[J]. IEEE transactions on systems, man, and cybernetics: systems, 2018, 48(9): 1622–1636.
- [18] ZHU Lin, CHUNG F L, WANG Shitong. Generalized fuzzy c-means clustering algorithm with improved fuzzy partitions[J]. IEEE transactions on systems, man, and cybernetics, part b, 2009, 39(3): 578–591.
- [19] 张远鹏, 邓赵红, 钟富礼, 等. 基于代表点评分策略的快速自适应聚类算法[J]. 计算机研究与发展, 2018, 55(1): 163–178.
- ZHANG Yuanpeng, DENG Zhaohong, ZHONG Fuli, et al. Fast self-adaptive clustering algorithm based on exemplar score strategy[J]. Journal of computer research and development, 2018, 55(1): 163–178.
- [20] DENG Zhaohong, JIANG Yizhang, CHUNG F L, et al. Transfer prototype-based fuzzy clustering[J]. IEEE transactions on fuzzy systems, 2016, 24(5): 1210–1232.
- [21] ASUNCION A, NEWMAN D J. UCI machine learning repository[EB/OL]. (2009-10-16)[2019-05-10] <http://archive.ics.uci.edu/ml>.
- [22] ZHOU Zhihua, LI Ming. Tri-training: exploiting unlabeled data using three classifiers[J]. IEEE transactions on knowledge and data engineering, 2005, 17(11): 1529–1541.
- [23] WANG Shuo, YAO Xin. Relationships between diversity of classification ensembles and single-class performance measures[J]. IEEE transactions on knowledge and data engineering, 2013, 25(1): 206–219.

作者简介:



张雄涛, 男, 1984 年生, 博士研究生, 主要研究方向为模式识别、模糊系统。



胡文军, 男, 1977 年生, 教授, 主要研究方向为模式识别、人工智能。



王士同, 男, 1964 年生, 教授, 主要研究方向为人工智能、数据挖掘、模糊系统。