

基于 AP 二次聚类的神经网络集成算法研究

李 辉^{1,3} 丁世飞^{1,2}

(中国矿业大学计算机科学与技术学院 徐州 221116)¹

(中国科学院智能信息处理重点实验室 北京 100190)²

(江苏师范大学计算机科学与技术学院 徐州 221116)³

摘 要 为了提高个体神经网络精度及差异度进而提高神经网络集成(Neural Network Ensemble, NNE)的泛化性能,提出了一种基于二次聚类的神经网络集成方法。首先对所有样本进行聚类,得到第一次聚类样本子集;然后对每一类样本子集进行二次聚类,得到每一子类的样本子集,通过 Affinity Propagation(AP)聚类使得“类内相似,类间相异”的准则最大化,类内样本能够反映真实的数据分布;最后按照排列组合的方式,从二次聚类的每个样本子集中选取一类样本构成训练集来构造一个个体神经网络。这样从不同类中选择样本集构造的个体神经网络差异性较大,既能使数据的规模较小,又能反映真实的数据分布,用这种方法产生的个体神经网络进行集成具有较高的性能。仿真实验表明,该方法能够取得较好的性能。

关键词 Affinity Propagation(AP)聚类,神经网络集成,二次聚类,个体神经网络

中图法分类号 TP18 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.2.046

AP Twice Clustering Based Neural Network Ensemble Algorithm

LI Hui^{1,3} DING Shi-fei^{1,2}

(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)¹

(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)²

(School of Computer Science and Technology, Jiangsu Normal University, Xuzhou 221116, China)³

Abstract In order to improve the precision and diversity of individual networks to improve the generalization performance of neural network ensemble (NNE), this paper proposed a method to generate individual neural network for ensembling based on twice clustering. Firstly, all the samples are chosen to cluster for the first time to form once clustering subclass, and then the twice clustering is performed for each type of subclass to form the sample subsets of each subclass. Affinity Propagation(AP) clustering makes the otherness criterion of “similar in classes, diversity between classes” maximization, and the samples in the class can response the real data distribution. Finally, according to the permutation and combination a subset is selected from each twice clustering of each subclass to construct a trainset to train an individual neural network. So the trainset with the smaller size of the data and the real data distribution can train the individual neural networks with the bigger diversity, and the ensemble of these individual neural networks can get better performance. Simulation experiments show that our proposed method here is effective.

Keywords Affinity propagation(AP) clustering, Neural network ensemble(NNE), Twice clustering, Individual neural network

1 引言

1990 年 Hansen 和 Salamon 证明:通过训练多个神经网络并将其结果进行集成,能显著地提高神经网络系统的泛化能力^[1]。由于神经网络集成(Neural Network Ensemble, NNE)在泛化能力上的优良性能以及精度的大幅提高,它自 1990 年被提出以来,得到了众多学者的广泛研究^[2-7]。NNE 由两部分组成,即个体生成部分和结论生成部分。文献[2,3]指出,集成学习有效的条件是每个单一的学习器错误率都应当低于 0.5,否则集成的结果反而会提高错误率。文献[4]给

出了神经网络集成泛化误差的分解公式,并表明只要个体神经网络泛化误差均值保持不变,增加差异性就可以提高网络的泛化能力,因此构造具有较大差异的神经网络在理论上被认为是集成方法所具有的重要特性。神经网络集成的泛化误差等于集成中个体网络的平均泛化误差和平均差异度之差^[5],因此要增强神经网络集成的泛化能力,一方面应尽可能提高个体网络的泛化能力,另一方面应尽可能地增大集成中各网络之间的差异。目前常用的神经网络个体生成方法是通过扰动训练数据来获得差异度较大的个体网络,主要有 Boosting^[8]类算法与 Bagging^[9]类算法两大类。Boosting 中各

到稿日期:2014-03-05 返修日期:2014-05-17 本文受国家重点基础研究计划(973 计划)项目(2013CB329502),国家自然科学基金(41074003,61379101)资助。

李 辉(1978-),男,博士,主要研究方向为人工智能、智能计算、神经网络等,E-mail:lihui400@163.com;丁世飞(1963-),男,教授,博士生导师,主要研究方向为机器学习、人工智能、智能计算、神经网络等。

网络的训练集决定于之前产生的网络的表现,被已有网络错误判断的示例将以较大的概率出现在新网络的训练集中。Boosting 算法有很多优点,如它有着较高的正确率、不需要先验知识、只需要选择合适的迭代次数等,但它的速度在一定程度上依赖于训练数据集和弱分类器的选择,训练数据不足或弱学习器太“弱”,都会导致训练精度的下降。Bagging 的基础是可重复取样,从原始训练集中随机抽取各神经网络的训练集,训练集的规模通常与原始训练集相当。这样,原始训练集中某些示例可能新的训练集中出现多次,而另外一些示例则可能一次也不出现,而且当数据规模较大时,训练数据集的规模也比较大。

聚类(Clustering)作为一种非监督学习方法,是智能计算领域中的一个重要的研究方向,同时,聚类技术也是数据挖掘中进行数据处理的重要分析工具和方法。AP(Affinity Propagation, AP)聚类算法是由 Frey B. J. 和 Dueck D. 于 2007 年提出的一种新的聚类算法,该算法无需事先定义类数,在迭代过程中不断搜索合适的聚类中心,自动从数据点间识别类中心(exemplars)的位置及个数。与其他聚类算法相比,AP 聚类算法具有如下优点:1)在处理大量数据时,时间快、精度高;2) AP 算法不需要事先指定聚类中心数目。因此,可以选用 AP 算法选取训练样本,通过 AP 聚类算法优化数据集后,将得到高质量、小样本的分类器训练样本集。

样本的分布情况对建立精确的网络模型是必要的,样本点的空间分布一般是不均匀的,有的区域样本点比较稠密,而有的区域样本点则比较稀疏。如果随机选取样本,就有可能导致各区域的抽样率不一样,得到样本之间的映射很难被网络学习,得到的网络也是不完备的。通过 AP 算法聚类样本,得到子空间样本,各子空间中样本间具有较大的相似度,能够反映样本的真实分布。分别对每个子空间中的样本进行二次聚类,得到每个子空间样本的子类,显然,各子空间的各类子类样本具有相近的数据分布,从各子空间的子类中分别选取一类样本构成一个训练集,既能使各训练集样本均匀分布,又能减小训练集的规模。

基于以上分析,本文结合聚类算法,提出了基于二次聚类的神经网络集成方法。首先对所有的样本进行聚类,得到第一次聚类样本子集;然后对每一类样本子集进行二次聚类,得到每一子类的样本子集,通过 AP 聚类使得“类内相似,类间相异”的准则最大化,类内样本能够反映真实的数据分布;最后按照排列组合的方式从二次聚类的每个样本子集中选取一类样本构成训练集来构造一个个体神经网络。这样从不同类中选择样本集构造的个体神经网络差异性较大,既能使数据的规模较小,又能反映真实的数据分布,用这种方法产生的个体神经网络进行集成具有较高的性能。本文第 2 节介绍了本文的相关理论基础:神经网络集成和聚类算法;第 3 节分析了基于二次聚类的神经网络集成方法,在此基础上给出了基于二次聚类的神经网络集成方法的体系结构、算法描述及分析;最后,通过仿真实验验证了本文提出的基于二次聚类的神经网络集成方法的有效性和性能。

2 相关理论

2.1 神经网络集成

神经网络集成是用有限个神经网络对同一个问题进行学习,集成在某输入示例下的输出,由构成神经网络集成的各种

神经网络在此示例下的输出共同决定^[4]。

假设输入 $x \in R^m$ 满足分布 $p(x)$,若在输入 x 下目标输出为 $d(x)$,神经网络 $f_i (i=1, 2, \dots, N)$ 的输出为 $f_i(x)$,则神经网络集成在输入 x 下的输出定义为:

$$f_{ensemble}(x) = \sum_{i=1}^N \omega_i f_i(x) \quad (1)$$

神经网络集成的泛化误差:

$$E_{ensemble} = \int P(x)(f_{ensemble}(x) - d(x))^2 dx \quad (2)$$

各神经网络 $f_i (i=1, 2, \dots, N)$ 的泛化误差:

$$E_i = \int P(x)(f_i(x) - d(x))^2 dx \quad (3)$$

各神经网络泛化误差的加权平均值为:

$$E_{average}(x) = \sum_{i=1}^N \omega_i E_i \quad (4)$$

神经网络 $f_i (i=1, 2, \dots, N)$ 的差异度:

$$A_i = \int P(x)(f_i(x) - d(x))^2 dx \quad (5)$$

神经网络集成的差异度:

$$A_{ensemble} = \sum_{i=1}^N \omega_i A_i \quad (6)$$

Krogh 等人通过理论分析得到如下神经网络集成泛化误差的计算公式:

$$E_{ensemble} = E_{average} - A_{ensemble} \quad (7)$$

由于各网络的差异度均非负,从式(7)可知,神经网络集成的泛化误差 $E_{ensemble}$ 不大于各神经网络的泛化误差的加权平均值 $E_{average}$,增大其差异度 $A_{ensemble}$ (也就是增大差异度 A_i) 能够有效地降低神经网络集成的泛化误差。

2.2 聚类算法

AP(Affinity Propagation, AP)聚类算法是由 Frey B. J. 和 Dueck D. 于 2007 年提出的一种新的聚类算法^[10,11]。该算法无需事先定义类数,同时把所有的数据点作为潜在的类中心,在迭代过程中不断搜索合适的聚类中心,自动从数据点间识别类中心(exemplars)的位置及个数^[12]。AP 算法以其简单、高效的优势已广泛应用于多种领域,如:设施选址^[13-15]、图像识别^[16]、图像分割^[17,18]、文本挖掘^[19]、生物医学^[10,20]、视频关键帧提取^[21]和图像检索^[22]等方面。王开军等人也对 AP 聚类算法进行了各种改进,并取得了较好的结果^[23-30]。与以前的方法不同,AP 算法^[31]不需要数据集具有某种特殊的结构,主要根据 N 个样本点之间的相似度进行聚类,这些相似度组成 $N \times N$ 的相似度矩阵 S ,如: $S(i, j)$ 表示样本点 i 和样本点 j 之间的相似度。矩阵 S 主对角线上的数值又称为 Preference,该值是对应的样本点能否成为聚类中心的评判标准,一般来说,其值越大表示这个点成为聚类中心的可能性就越大。AP 算法主要依靠一种“信息传递”机制实现数据集的聚类。这种消息传递机制中主要包含两类信息: Responsibility 和 Availability。Responsibility 表示样本点对不同的候选类中心发出的信息,表明候选类中心相应于该样本点作为潜在类中心的适合程度,该值越大表明候选类中心越可能成为实际的类中心; Availability 表示候选类中心对样本点发出的信息,表明该样本点相应于候选类中心的聚合程度,该值越大表明样本点越可能属于某一类。

3 基于聚类的神经网络集成算法

3.1 AP 聚类算法

AP 算法通过迭代过程不断更新每一个点的 Responsibility 和

Availability 值,直到自动产生若干个类中心,同时将其余的数据点分配到相应的类团中。AP 算法的步骤如下:

1) 算法初始化,计算初始相似度矩阵 S ,对 $Preference$ (P) 赋初值。

2) 计算样本点间的 Responsibility 值:

$$R(i, k) \leftarrow s(i, k) - \max_{j \neq k} (s(i, j) + A(i, j))$$

$A(i, j)$ 表示 j 对于 i 的 Availability 值。

3) 计算样本点间的 Availability 值:

$$A(i, k) \leftarrow \min\{0, R(k, k) + \sum_{j \neq i, k} \max(0, R(j, k))\}$$

$$A(k, k) \leftarrow \sum_{j \neq k} \max(0, R(j, k))$$

4) Responsibility 和 Availability 的更新:

$$R_{t+1}(i, k) = \lambda \cdot R_t(i, k) + (1 - \lambda) \cdot R_{t+1}^{old}(i, k)$$

$$A_{t+1}(i, k) = \lambda \cdot A_t(i, k) + (1 - \lambda) \cdot A_{t+1}^{old}(i, k), \lambda \in [0.5, 1)$$

$$A_{t+1}(k, k) = P(k) - \max_{j \neq k} [A_{t+1}(k, j) + S_{t+1}(k, j)],$$

$$j \in \{1, 2, \dots, N\}, j \neq k$$

其中, λ 是收敛系数,主要用于调节算法的收敛速度及迭代过程中的稳定性。

5) 如果迭代次数超过设定的最大值或者当聚类中心在若干次迭代中不发生改变时,则终止计算,确定类中心及各类的样本点;否则返回 2),继续计算。

6) 根据 AP 聚类结果,记录类中心及其类样本。

研究表明,使用收缩因子(Constriction Factor)可以有效地保证算法收敛^[11],收缩因子的定义如式(8)所示。

$$\rho = \frac{2}{|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}|}, \varphi > 4 \quad (8)$$

为加速算法收敛,对 Responsibility 和 Availability 进行更新,如式(9)、式(10)所示。

$$R_{t+1}(i, k) = \rho \cdot \lambda \cdot R_t(i, k) + (1 - \lambda) \cdot R_{t+1}^{old}(i, k) \quad (9)$$

$$A_{t+1}(i, k) = \rho \cdot \lambda \cdot A_t(i, k) + (1 - \lambda) \cdot A_{t+1}^{old}(i, k), \quad \lambda \in [0.5, 1) \quad (10)$$

3.2 基于聚类的神经网络集成算法描述

与其它聚类算法不同,AP 聚类算法无需事先定义类数和类中心,同时把所有的数据点作为潜在的类中心,在迭代过程中不断搜索合适的聚类中心,自动从数据点间识别类中心(exemplars)的位置及个数^[12]。因此选择 AP 算法作为本文的聚类算法。如下所述,通过二次聚类,从每一类二次聚类的聚类样本子集中选择一类样本构成一个训练集用于训练一个个体神经网络,这样可以满足:1)减小训练集的数据规模,达到数据降维的目的;2)使得“类内相似,类间相异”的聚类准则最大化,得到的样本集能够反映真实的数据分布;3)从不同二次聚类结果中选择子样本构成训练集能够构造出较大差异度的个体神经网络。具体基于二次聚类的神经网络集成方法过程分析如下:

首先对所有样本进行聚类,得到第一次聚类样本子集;然后对每一类样本子集进行二次聚类,得到每一子类的样本子集,通过 AP 聚类使得“类内相似,类间相异”的准则最大化,类内样本能够反映真实的数据分布;最后按照排列组合的方式从二次聚类的每个样本子集中选取一类样本构成训练集来构造一个个体神经网络。这样从不同类中选择样本集构造的个体神经网络差异性较大,既能使数据的规模较小,又能反映真实的数据分布,用这种方法产生的个体神经网络进行集成具有较高的性能。图 1 给出了基于二次聚类的神经网络集成

方法的体系结构。

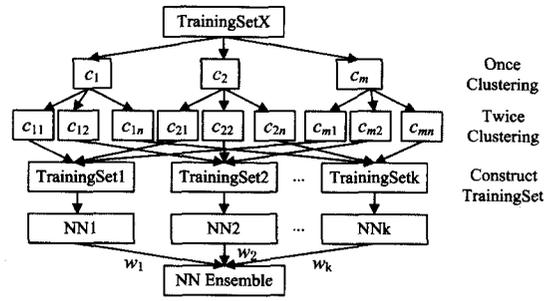


图 1 基于二次聚类的神经网络集成方法体系结构

相应的算法描述如下:

Step1 训练样本集 $X = \{x_1, x_2, \dots, x_N\}$,用 AP 算法第一次聚类样本,记录类中心 $c_i (i = 1, 2, \dots, m)$ 及类内数据, c_i 为一次聚类样本类数, $\|c_i\|$ 为样本的数目,其中 $c_i = \{x_i^1, x_i^2, \dots, x_i^{\|c_i\|}\}$, $X = c_1 \cup c_2 \cup \dots \cup c_m$;

Step2 用 AP 算法对 $c_i (i = 1, 2, \dots, m)$ 进行二次聚类,记录类中心 $c_{ij} (i = 1, 2, \dots, m; j = 1, 2, \dots, n)$ 及类内数据, c_{ij} 为对一次聚类中第 c_i 类样本二次聚类的样本类数, $\|c_{ij}\|$ 为样本的数目,其中 $c_{ij} = \{x_{ij}^1, x_{ij}^2, \dots, x_{ij}^{\|c_{ij}\|}\}$;

Step3 用一定的规则分别从每个 $c_i (i = 1, 2, \dots, m)$ 样本的二次聚类结果中抽取一个样本,组成新的训练样本,产生 k 个训练子集,利用这些训练子集训练神经网络,得到对应的 k 个个体神经网络。

Step4 集成 k 个个体神经网络输出结果。

3.3 集成个体神经网络结果

结论生成方法通常由各网络的输出通过简单平均(等权)^[6,7,32,33]或加权平均产生^[34]。在 Bagging 中采用简单平均,而 Boosting 使用加权平均。关于是否使用权重,目前有不同的观点。Perrone 等人^[35,36]认为采用加权平均法比采用简单平均法得到更好的泛化能力,并给出了指导权值选择的公式,但也有研究者^[37]认为,对权值的优化过程会导致过拟合,从而使集成的泛化能力降低,因此提倡使用简单平均。本文提出的基于二次聚类的神经网络集成算法侧重于个体网络生成方面,通过提高个体神经网络精度和泛化性能,达到网络集成的目的,得到所希望的网络集成效果,因此集成的输出采用简单平均(等权)的方法。

4 仿真实验

本节提供的实验数据来自 UCI 标准数据库,用 Iris、Wine 和 Zoo 3 种数据库验证了本文提出的基于 AP 二次聚类的神经网络集成算法的有效性。表 1 列出了 Iris、Wine 和 Zoo 3 种数据库的类数、样本数和属性数。

表 1 3 种数据集参数

数据集	类数	样本数	属性数
Iris	3	150	4
Wine	3	178	13
Zoo	7	214	16

首先对所有的样本进行聚类,得到第一次聚类样本子集 $c_i (i = 1, 2, \dots, m)$ 。图 2 给出了在 Iris 数据集上一次聚类的迭代过程。图 3 给出了在 Iris 数据集上一次聚类的聚类结果。然后对每一类样本子集 $c_i (i = 1, 2, \dots, m)$ 进行二次聚类,

得到每一子类的样本子集 c_{ij} ($i=1, 2, \dots, m; j=1, 2, \dots, n$)。图4—图9给出了在 Iris 数据集上二次聚类的迭代过程及其相应的聚类结果,在其它数据集上二次聚类的迭代过程及其相应的聚类结果在此不再赘述。在各种神经网络模型中, BP (Back Propagation, BP) 具有较强的非线性模拟能力, 是用得最多的一种模型, 因此, 本文选用 BP 网络作为待集成的个体神经网络。根据个体神经网络输入和输出神经元的个数确定个体神经网络隐层神经元的个数为 2—13。表 2 给出了个体神经网络隐层神经元数量对个体神经网络性能的影响, 本文选择个体神经网络误差率达到最小时的网络隐层神经元数。

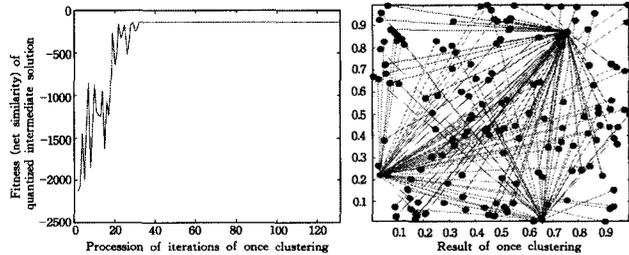


图2 一次聚类迭代过程

图3 一次聚类结果

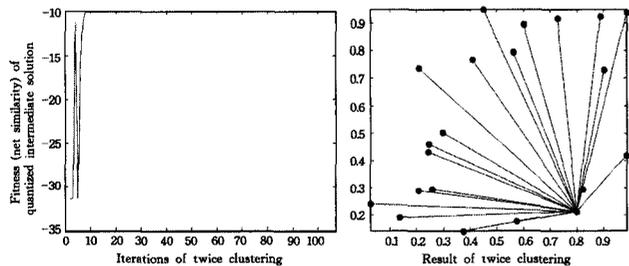


图4 对 c_1 的二次聚类迭代过程

图5 对 c_1 的二次聚类结果

表2 个体神经网络隐层神经元数量对个体神经网络性能的影响

数据集	个体神经网络取不同神经元的误差率 (%)											
	2	3	4	5	6	7	8	9	10	11	12	13
Iris	0.074	0.032	0.006	0.001	0.0010	0.0035	0.0025	0.0009	0.0010	0.0010	0.0009	0.0008
Wine	0.082	0.031	0.058	0.003	0.0003	0.0045	0.0035	0.0008	0.0009	0.0009	0.0010	0.0009
Zoo	0.084	0.035	0.061	0.005	0.0008	0.0010	0.0027	0.0007	0.0008	0.0010	0.0010	0.0010

表3 个体神经网络及神经网络集成的误差率

方法	误差率 (%)		
	Iris	Wine	Zoo
个体神经网络 1	0.84e-4	0.69e-4	0.72e-4
个体神经网络 2	0.63e-4	0.74e-4	0.68e-4
个体神经网络 3	0.46e-4	0.70e-4	0.71e-4
个体神经网络 4	0.52e-4	0.72e-4	0.74e-4
个体神经网络 5	0.44e-4	0.71e-4	0.73e-4
直接分类	0.96e-4	0.95e-4	0.97e-4
神经网络集成	0.58e-4	0.71e-4	0.72e-4

结束语 本文结合聚类算法, 提出了基于二次聚类的神经网络集成方法。首先对所有样本进行聚类, 得到第一次聚类样本子集; 然后对每一类样本子集进行二次聚类, 得到每一子类的样本子集, 通过 AP 聚类使得“类内相似, 类间相异”的准则最大化, 类内样本能够反映真实的数据分布; 最后按照排列组合的方式从二次聚类的每个样本子集中选取一类样本构成训练集构造一个个体神经网络。这样从不同类中选择样本集构造的个体神经网络差异性较大, 既能使数据的规模较小, 又能反映真实的数据分布, 对这种方法产生的个体神经网络进行集成具有较高的性能。仿真实验表明, 本文提出基于二次聚类的神经网络集成方法能够获得更高的性能。下一步我

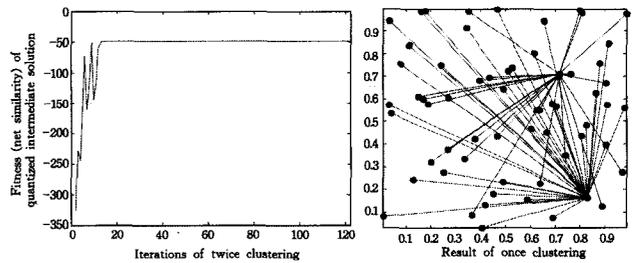


图6 对 c_2 的二次聚类迭代过程

图7 对 c_2 的二次聚类结果

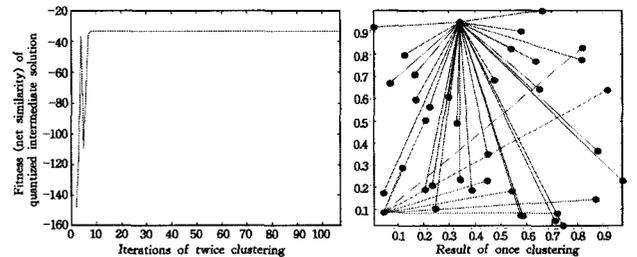


图8 对 c_3 的二次聚类迭代过程

图9 对 c_3 的二次聚类结果

表 3 分别比较了个体神经网络及其集成的误差率(由于本文实验在 Iris、Wine 和 Zoo 等小数据集上进行, 个体神经网络的个数根据排列组合出的训练集个数确定)。总的来说, 本文提出的基于二次聚类的神经网络集成算法在小数据集上具有更好的精度, 从集成的结果来看, 个体神经网络具有较大的差异度, 获得了较好的性能。因此, 通过二次聚类方法构造的样本子集能够反映真实的数据分布, 而且使数据的规模较小, 用这些样本构成训练集能够训练出精度和差异度较大的个体神经网络。

们将探索个体神经网络及其隐层神经元个数的自适应确定方法, 并将基于二次聚类的神经网络集成方法应用于现实生活, 以进一步验证该方法的实际应用价值。

参考文献

- [1] Hansen L K, Salamon P. Neural network ensemble [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1990, 12 (10):993-1001
- [2] Valentini G, Masulli F. Ensembles of learning machines [C] // Lecture Notes in Computer Sciences. Heidelberg, Germany: SpringerVerlag, 2002
- [3] Yu Shi-xin. Feature Selection and Classifier Ensembles: A study on hyperspectral remote sensing data [OL]. 2003. http://143.129.203.3/visielab/theses/shixin/thesis_yu.pdf
- [4] Tesauo G, Touretzky D S, Leen T K, et al. Neural network ensembles, Crosses validation, and active learning [C] // Advances in Neural Information Processing System. Cambridge, MA: MIT Press, 1995: 231-238
- [5] Zhou Zhi-hua, Chen Shi-fu. Neural network ensemble [J]. Chinese Journal of Computers, 2002, 25(1): 1-8

(下转第 252 页)

- [17] 《运筹学》教材编写组. 运筹学(第四版)[M]. 北京:清华大学出版社,2012
- [18] Moore R E. A test for existence of solutions to nonlinear systems[J]. SIAM Journal on Numerical Analysis, 1977, 14(4): 611-615
-
- (上接第 227 页)
- [6] Li Hui, Ding Shi-fei. Research of Individual Neural Network Generation and Ensemble Algorithm Based on Quotient Space Granularity Clustering[J]. Applied Mathematics & Information Sciences, 2013, 7(2): 701-708
- [7] Xu Xin-zheng, Ding Shi-fei, Jia Wei-kuan, et al. Research of assembling optimized classification algorithm by neural network based on Ordinary Least Squares (OLS) [J]. Neural Comput & Applic, 2013(22): 187-193
- [8] Schapire R E. The strength of weak learnability[J]. Machine Learning, 1990, 5(2): 197-227
- [9] Breiman L. Bagging predictors[J]. Machine Learning, 1996, 24(2): 123-140
- [10] Frey B J, Dueck D. Clustering by passing messages between data points[J]. Science, 2007, 315(5814): 972-976
- [11] Liu Xiao-yong, Fu Hui. A fast affinity propagation clustering algorithm[J]. Journal of Shandong University (Engineering Science), 2011, 41(4): 20-28
- [12] Xia Ding-yin, Wu Fei, Zhang Xu-qing, et al. Local and global approaches of affinity propagation clustering for large scale data [J]. Journal of Zhejiang University-Science A, 2008, 9(10): 1373-1381
- [13] Lazić N, Givoni I, Aarabi P, et al. FLoSS: Facility location for subspace segmentation [C] // Proceedings of 12th International Conference on Computer Vision (ICCV). Kyoto: IEEE Press, 2009: 825-832
- [14] Lazić N, Frey B J, Aarabi P. Solving the uncapacitated facility location problem using message passing algorithms [C] // Proceedings of 13th International Conference on Artificial Intelligence and Statistics (AISTATS). Sardinia: Microtome Publishing, 2010: 429-436
- [15] Tang Dong-ming, Zhu Qing-xin, Yang Fan, et al. Solving large-scale location problem using affinity propagation clustering [J]. Application Research of Computers, 2010, 27(3): 841-844
- [16] Dueck D, Frey B J. Nonmetric affinity propagation for unsupervised image categorization [C] // Proceedings of 11th International Conference on Computer Vision (ICCV). Rio de Janeiro: IEEE Press, 2007: 1-8
- [17] Zhang Ren-yan, Zhao Hong-liang, Lu Xiao, et al. Grey image segmentation method based on affinity propagation clustering [J]. Journal of Naval University of Engineering, 2009, 21(3): 33-37
- [18] Givoni I E, Frey B J. Semi-supervised affinity propagation with instance-level constraints [C] // Proceedings of 12th International Conference on Artificial Intelligence and Statistics (AISTATS). Florida: Microtome Publishing, 2009: 161-168
- [19] Guan Ren-chu, Pei Zhi-li, Shi Xiao-hu, et al. Weight affinity propagation and its application to text clustering [J]. Journal of Computer Research and Development, 2010, 47(10): 1733-1740
- [20] Dueck D, Frey B J, Jojic N, et al. Constructing treatment portfolios using affinity propagation [C] // Proceedings of International Conference on Research in Computational Molecular Biology (RECOMB). Singapore: Springer, 2008: 360-371
- [21] Xu Wen-zhu, Xu Li-hong. Adaptive key-frame extraction based on affinity propagation clustering [J]. Computer Science, 2010(1): 268-270
- [22] Xiang Pei-su. New CBIR system based on the affinity propagation clustering algorithm [J]. Journal of Southwest University for Nationalities: Natural Science Edition, 2010, 36(4): 624-627
- [23] Wang Kai-jun, Zhang Jun-ying, Li Dan, et al. Adaptive affinity propagation clustering [J]. Acta Automatica Sinica, 2007, 33(12): 1242-1246
- [24] Wang Kai-jun, Li Jian, Zhang Jun-ying, et al. Semi-supervised affinity propagation clustering [J]. Computer Engineering, 2007, 33(23): 197-198, 201
- [25] Wang Kai-jun, Zheng Jie. Fast algorithm of affinity propagation clustering under given number of clusters [J]. Computer Systems & Applications, 2010, 19(7): 207-209
- [26] Xie Xin-xi, Wang Shi-tong. Affinity propagation clustering for symbolic interval data based on mutual distance [J]. Computer Application, 2008, 28(6): 1441-1443
- [27] Xiao Yu, Yu Jian. Semi-Supervised clustering based on affinity propagation algorithm [J]. Journal of Software, 2008, 19(11): 2803-2813
- [28] Gu Rui-jun, Wang Jia-cai, Chen Geng, et al. Affinity propagation clustering for large scale dataset [J]. Computer Engineering, 2010, 36(23): 22-24
- [29] Dong Jun, Wang Suo-ping, Xiong Fan-lun. Affinity propagation clustering based on variable-similarity measure [J]. Journal of Electronics & Information Technology, 2010, 32(3): 509-514
- [30] Li Ya-qin, Yang Hui-zhong. Multi-model modeling method based on affinity propagation clustering and Gaussian processes [J]. Computers and Applied Chemistry, 2010, 27(1): 51-54
- [31] Frey B J, Dueck D. Clustering by passing messages between data points [J]. Science, 2007, 315(5814): 972-976
- [32] Lin J, Zhu B Z. Improved Principal Component Analysis and Neural Network Ensemble Based Economic Forecasting [J]. Lecture Notes in Computer Science, 2006, 4113: 135-145
- [33] Lin Jian, Peng Min-jing. GDP forecasting model based on neural networks ensemble [J]. Chinese Journal of Management, 2005, 2(4): 434-436
- [34] Lin Jian, Zhu Bang-zhu. Neural network ensemble based on forecasting effective measure and its application [J]. Journal of Computational Information Systems, 2005, 1(4): 781-787
- [35] Perrone M P, Cooper L N. When Networks Disagree: Ensemble method for neural networks [C] // Mammon R J, eds. Artificial Neural Networks for Speed and Vision. New Yorks: Chapman & Hall, 1993: 126-142
- [36] Li Ling-ling, Liu Xi-yu, Lu Shu-qiang. Constructive methods for parallel learning neural network ensemble based on particle swarm optimization [J]. ShanDong Science, 2007, 20(4): 16-20
- [37] Opitz D, Vedelsby J. Actively searching for an efficient neural network ensemble [J]. Connection Science, 1996, 8(3/4): 337-353