

# 基于多学位识别的学位决策机制的研究与实现

刘 敏

(西华师范大学计算机学院 南充 637002)

**摘 要** 在以 1:1 数字学习方式为主的自主学习模式下存在相关专业多学位识别的问题。针对该问题,建立了一种使用遗传算法和 BP 神经网络的多学位识别机制。该机制根据问题的特点,采用遗传算法产生样本群体,并用遗传算法确定神经网络模型的参数,通过神经网络自适应学习和训练,找出输入和输出的关系,从而达到多学位识别的目的。实验验证了该方法的有效性。

**关键词** 遗传算法, BP 神经网络, 数字学习, 多学位识别

## Research and Implementation of Decision-making Mechanism Based on Identification of Multi-degree

LIU Min

(Computer Institute, China West Normal University, Nanchong 637002, China)

**Abstract** The genetic algorithm optimization back-propagation BP neural network decision-making mechanism was presented in order to overcome the shortcoming of the independent study model based on the digital technology supported learning that can't identify a number of degrees. According to its feature, the genetic algorithm is adopted to produce sample groups and determine parameters of the neural network model. The relationship between input and output has been identified during adaptive learning and training of the neural network, so as to achieve the purpose of identification of multi-degree. It validates the proposed approach by experiments.

**Keywords** Genetic algorithm, BP neural network, Digital technology supported learning, Identification of multi-degree

## 1 前言

网络技术的发展使 Internet 网不断向下一代互联网迈进。不管网络协议如何,下一代互联网的实现所要面临的主要挑战除网络的平滑过渡以外,还有网络应用与业务的创新。远程教育在网络传统应用之一,它的创新在于构筑基于下一代移动网络的以 1:1 数字学习(Digital Technology Supported Learning)<sup>[1]</sup>方式为主的自主学习新模式。在自主学习模式下存在专业学位决策的问题,虽然使用数据挖掘算法能够准确地得出一个唯一的学位专业<sup>[2]</sup>,但对有些极端情况并不适用。如某些自主学习者可能同时完成了多个相关专业的学习并都已达到授位学分,这种情况下如果仍然只能识别出一个专业学位,那么对自学者而言就是一种利益损害。为了正确识别学习者相关专业多学位的情况,从而为高校管理决策者做出正确决策提供准确的依据,本文提出了一种基于遗传算法(Genetic Algorithm, GA)和 BP 神经网络的多学位识别机制。

神经网络作为知识获取的工具,已被设计成仿真人脑获取知识的一种组织。BP 神经网络通过输入当前的学习样本数据,然后调整网络中的权重来完成自身的学习。神经网络不但可以学习、积累并储存给定的知识,而且可以在环境变化时进行重新学习<sup>[3]</sup>。由于人工神经网络的分类、识别和记忆能力不仅取决于神经元模型和网络拓扑结构,而且取决于网

络学习算法,所以本文通过遗传算法产生样本群体,并用遗传算法确定神经网络模型的参数。

遗传算法是模拟自然界遗传机制和生物进化论而成的一种并行随机搜索最优化方法<sup>[4]</sup>。它将“优胜劣汰,适者生存”的生物进化原理引入到优化参数形成的编码串群体中,按所选择的适配值函数通过遗传中的复制、交叉及变异对个体进行筛选,使适配值高的个体被保留下来,组成新的群体,新的群体既继承了上一代的信息,又优于上一代。这样周而复始,群体中个体适应度不断提高,直到满足一定的条件<sup>[5]</sup>。遗传算法由于适用条件非常宽松,不需诸如连续、可导和单峰等要求,因而适用于各种场合<sup>[6,7]</sup>。

## 2 影响学位专业名称形成的要素

学位的授予建立在严格的培养和考核的基础之上,下一代互联网的自主学习者在学习过程中具有很强的学习自主权,倾向于哪个学位,是两个还是多个,完全取决于其是否达到某一学位的授予标准。纵观国内各高校现在全部实行的是学分制,简单说来学分制的学位授予标准就是“学习完一门课程,通过这门课程的考核即可取得该门课程的学分,当学分总值达到授位标准时则授予相应专业学位”。由此可见,影响学位专业名称形成的要素有两个:一是是否通过一门课程的考核,二是取得的专业课程(包括专业选修课)学分总数是否达到该专业授位标准要求的总学分。

### 3 本文的识别机制

#### 3.1 神经网络模型

本文实现多学位识别学位决策机制的神经网络模型如图1所示。在该模型中,需要确定以下参数:

- $m$  为神经网络隐含层神经元的数目;
- $WI$  为神经网络的输入层权重;
- $TI$  为神经网络的输入层阈值;
- $TFI$  为神经网络的输入层转移函数;
- $WH$  为神经网络的隐含层权重;
- $TH$  为神经网络的隐含层阈值;
- $NTF$  为神经网络的网络训练函数。

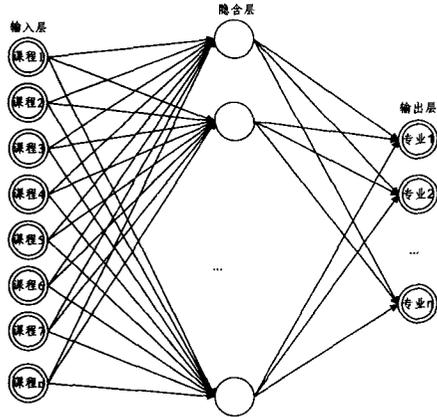


图1 多学位识别的神经网络模型

#### 3.2 采用遗传算法确定神经网络模型参数

##### 3.2.1 种群初始化

①根据经验公式  $m = \text{fix}[\log_2 n] + 1$  来确定神经网络隐含层神经元个数,这里,  $m$  表示隐含层神经元个数;  $n$  表示学习样本数目;  $\text{fix}(x)$  表示取整操作。

②根据式(1)将神经网络的参数空间  $[L, U]$  分割成  $B$  个子空间。

$$\begin{cases} L_i = L + (i-1)[(u_k - l_k)/B]l_k \\ U_i = U + (B-i)[(u_k - l_k)/B]l_k \end{cases} \quad i=1, 2, \dots, B$$

$$u_k - l_k = \max_{1 \leq i \leq n} \{u_i - l_i\} \quad (1)$$

其中,  $L = [l_1, l_2, \dots, l_n]^T$  和  $U = [u_1, u_2, \dots, u_n]^T$  分别表示  $n$  个自变量(神经网络的参数)的下边界和上边界;  $B$  为设计参数;  $l_k$  表示第  $k$  位为1,其它位为0的  $n$  维列向量;  $L_i$  和  $U_i$  分别表示类似于  $L$  和  $U$  的  $n$  维列向量;经过本操作,最终将神经网络的参数空间  $[L, U]$  分割成  $[L_1, U_1], [L_2, U_2], \dots, [L_B, U_B]$  这样的  $B$  个子空间。

③按照式(2)将每个子空间进行离散化。假设自变量  $x_i$  的定义域为  $[l_i, u_i]$ ,根据设计参数  $Q_1$  ( $Q_1$  是奇数),将自变量  $x_i$  量化成  $Q_1$  个水平  $a_{i1}, a_{i2}, \dots, a_{iQ_1}$ ,具体的  $a_{ij}$  计算公式如下:

$$a_{ij} = \begin{cases} l_i & j=1 \\ l_i + (j-1) \left[ \frac{(u_i - l_i)}{(Q_1 - 1)} \right] & 2 \leq j \leq Q_1 \\ u_i & j=Q_1 \end{cases} \quad (2)$$

④从每个子空间中挑选出  $M_1$  个染色体。首先构造正交表  $L_{M_1}(Q_1^N) = [a_{ij}]_{M_1 \times N}$ ,此处的  $N$  为神经网络参数空间的维数,  $M_1 = Q_1^{J_1}$ ,  $J_1$  是满足条件  $\frac{Q_1^{J_1} - 1}{Q_1 - 1} \geq N$  的最小正整数;然

后从这  $Q_1^N$  个组合中选取  $M_1$  个组合;最后应用这  $M_1$  个组合生成  $M_1$  个染色体。

⑤根据适应度值的优劣,从上面已产生的  $M_1 B$  个潜在的染色体中,选择其中最优的  $G$  个染色体为初始种群,这里的  $G$  是初始种群的大小。

##### 3.2.2 适应度评价

对于当前种群中的单个个体,也就是神经网络的一组参数,采用该神经网络的误差来度量这个个体的适应度值。本文的适应度评价函数如下所示:

$$\text{FITNESS}(pop) = \text{MSE}(Y_L, \hat{Y}) = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^k \{\omega_j [Y_L(i, j) - \hat{Y}(i, j)] / Y_L(i, j)\}^2}{n-1}} \quad (3)$$

其中,  $pop$  表示当前种群中的单个个体(神经网络的一组参数);  $Y_L$  表示学习样本的给定输出值;  $\hat{Y}$  表示采用给定神经网络计算得到的学习样本的计算输出值;  $n$  表示学习样本的个数,  $k$  表示给定仿真优化问题绩效指标的个数,  $\omega_j$  表示第  $j$  个绩效指标的权重,  $Y_L(i, j)$  表示第  $i$  个学习样本的第  $j$  个绩效指标的给定值,  $\hat{Y}(i, j)$  表示采用给定神经网络计算得到的第  $i$  个学习样本的第  $j$  个绩效指标的计算值。

##### 3.2.3 交叉操作

①按照交叉概率选择进行交叉操作的两个父代染色体;假设要进行交叉操作的两个父代个体为:

$$\begin{cases} p_1 = (p_{11}, p_{12}, \dots, p_{1N}) \\ p_2 = (p_{21}, p_{22}, \dots, p_{2N}) \end{cases}$$

则定义它们的求解空间  $[l_{parent}, u_{parent}]$  为:

$$\begin{cases} l_{parent} = [\min(p_{11}, p_{21}), \min(p_{12}, p_{22}), \dots, \min(p_{1N}, p_{2N})] \\ u_{parent} = [\max(p_{11}, p_{21}), \max(p_{12}, p_{22}), \dots, \max(p_{1N}, p_{2N})] \end{cases}$$

②将要进行交叉操作的两个父代个体的求解空间离散化。将要进行交叉操作的两个父代个体的求解空间  $[l_{parent}, u_{parent}]$  离散化成  $Q_2$  份(这里的  $Q_2$  为设计参数),具体的离散化方法同式(2)。

③将所有自变量进行分组。为了避免在选择过程中进行大规模的种群点评估,每对父代尽可能不要产生太多的潜在子代点。鉴于此,将变量  $x_1, x_2, \dots, x_N$  分成  $F$  组,每一组变量将会被看成一个因素,这里的  $F$  是一个很小的设计参数。这样,相应正交表的组合数目将会减少,同时也就产生一个规模比较小的初始种群。这里,随机地生成  $F-1$  个整数  $k_1, k_2, \dots, k_{F-1}$ ,假定  $1 < k_1 < k_2 < \dots < k_{F-1} < N$ ,对于每个染色体  $x = (x_1, x_2, \dots, x_N)$  产生以下  $F$  个因素。

$$\begin{cases} f_1 = (x_1, \dots, x_{k_1}) \\ f_2 = (x_{k_1+1}, \dots, x_{k_2}) \\ \dots \\ f_N = (x_{k_{F-1}+1}, \dots, x_N) \end{cases} \quad (4)$$

④应用正交表从父代求解空间中选择潜在子代点。首先生成正交表  $L_{M_2}(Q_2^F) = [b_{ij}]_{M_2 \times F}$ ,此处的  $Q_2$  是一个奇数,  $M_2 = Q_2^{J_2}$ ,  $J_2$  是满足条件  $\frac{Q_2^{J_2} - 1}{Q_2 - 1} \geq F$  的最小正整数;然后从这  $Q_2^F$  个组合中选取  $M_2$  个组合;最后应用这  $M_2$  个组合生成  $M_2$  个潜在子代。

⑤从这  $M_2 Q_2$  个潜在子代个体和两个父代个体中,选择

适应度值最好的两个个体,作为本次交叉操作的结果。

⑥如果当前已进行的交叉操作次数已达到预设值,则停止交叉操作;否则,转至步骤①。

### 3.2.4 变异操作

①按照变异概率随机选择一个需要进行变异操作的父代染色体;

②随机选择需要进行变异操作的某个基因位(自变量);

③按照微摄动方法,得到变异后的子代染色体;这里的微摄动方法是指,将父代染色体中的已选基因分别微调为原来的 $1-2\sigma, 1-\sigma, 1+\sigma$ 和 $1+2\sigma$ ,这样就得到变异后的4个子代染色体。这里的 $\sigma$ 为一个设计参数;

④从父代染色体和子代染色体中选择一个最优个体作为此次变异的结果;

⑤如果当前已进行的变异操作次数已达到预设值,则停止变异操作;否则,转至步骤①。

### 3.3 多学位识别的神经网络算法思想

①按上述第3.2节采用的遗传算法初始化网络参数。

②输入 $P_i$ (由遗传算法得到的样本值)作为输入值, $P_i = \{p_{i1}, p_{i2}, \dots, p_{ij}\}$ ,  $p_{ij}$ 表示第 $i$ 个学生第 $j$ 门课程的学习情况,考核通过则 $p_{ij} = 1$ ,未学或考核未通过则 $p_{ij} = 0$ 。其中, $1 \leq j \leq$ 所有专业课程总数。以西华师范大学计算机学院为例,该学院共有计算机科学、软件工程和通信工程3个专业,都属于计算机相关专业,其中计算机科学有30门专业课、软件工程专业32门、通信工程专业34门,去掉交叉课程的重复数,如:计算机应用基础、数据结构、微机原理等课程,所有专业课程总数统计为83,即 $1 \leq j \leq 83$ 。

③求隐含层神经元的输入及输出:

$$in_j = \sum WI_{ij} \times P_i + TI_j$$

$$out_j = (1 + e^{-in_j})^{-2}$$

其中, $in_j, out_j$ 分别表示隐含层第 $j$ 个神经元的输入与输出。

④求输出层神经元的输入与输出:

$$in_j2 = \sum WH_{jk} \times out_j + TH_k$$

$$out_j2 = (1 + e^{-in_j2})^{-2}$$

⑤计算误差:

$$d_k = (out_j2 - T_k) \times out_j2 \times (1 - out_j2)$$

$$\delta_j = \sum d_k \times WH_{jk} \times out_j \times (1 - out_j)$$

其中, $d_k$ 为输出层第 $k$ 个神经元的一般化误差, $\delta_j$ 为隐含层第 $j$ 个神经元相对每个神经元的误差, $T_k$ 为期望输出值。

⑥修改 $WI_{ij}, WH_{jk}, TI_j, TH_k$ 的值:

$$WH_{jk} = WH_{jk} + \alpha \times d_k \times in_j$$

$$TH_k = TH_k + \beta \times d_k$$

$$WI_{ij} = WI_{ij} + \alpha \times \delta_j \times p_{im}$$

$$TI_j = TI_j + \beta \times \delta_j \alpha$$

$$E_c = \sum E_l = \sum 0.6 \times (out_j - T_k)^2$$

其中 $\alpha, \beta$ 为常数,分别为学习效率 and 系数,并且 $0 < \alpha < 1, 0 < \beta < 1$ 。

⑦如果有下一个学习样本,则输入并转向③;否则转向⑧。

⑧计算误差函数:

$$E_c = \sum E_l = \sum 0.6 \times (out_j - T_k)^2, l=1, 2, \dots, N$$

其中, $E_l$ 为第 $L$ 个模式的误差函数值, $L=1, 2, \dots, N, N$ 为学习样本个数, $E$ 为整个模式集的误差函数值。如果 $E$ 小于 $E_0$ ( $E_0$ 为预先给定的总误差判别标准值), $E_l$ 小于 $E_c$ ( $l=1, 2, \dots, N$ ),或 $E_c$ 为预先给定的第 $L$ 个学习样本误差判别标准值,则转向⑨;否则重复步骤③—步骤⑥。

⑨将 $WI_{ij}, WH_{jk}, TI_j, TH_k$ 的值作为知识保存起来,用以对某个学习者进行学位专业名称的多学位识别。

## 4 实验与结论

以西华师范大学计算机学院的计算机科学(A)、软件工程(B)和通信工程(C)3个专业(均为计算机相关专业)为例,分别选取单学位样本3个(A、B、C样本各1个)、双学位样本3个("A+B"、"A+C"、"B+C"样本各1个)、三学位样本1个("A+B+C"样本)作为初始群体,通过选择、交叉及变异等操作产生出具有300个样本的新群体,其中的280个作为学习样本来确定本文决策机制中的相关参数,剩余的20个则是测试样本。通过实验,发现本文方法的识别误差都在6%以内,平均识别误差为2.471%。实验表明,本文方法的识别结果非常理想,尤其是针对那些有较多交叉课程的相关专业的多学位识别,使用遗传算法和BP神经网络的多学位识别机制是可行、正确和有效的。当然,由于以1:1数字学习方式为主的自主学习模式还未完全形成,因此本文的方法还有待在将来的实践中进一步得到验证和完善。

## 参考文献

- [1] ChanTak-wai. One-to-one Technology-enhanced Learning: an Opportunity for Global Research Collaboration[J]. Research and Practice in Technology Enhanced Learning, 2006, 1(1): 3-29
- [2] 刘敏,滕华,毛嘉莉,等.基于自主学习环境下使用C4.5算法的专业预测[J].计算机系统应用, 2008(3): 38-41
- [3] Kryzanowski L, Galler M, Wright D W. Using Artificial Neural Networks to Pick Stocks[J]. Financial Analysts Journal, 1993, 49(2): 21-27
- [4] 杨建刚. 人工神经网络实用教程[M]. 杭州: 浙江大学出版社, 2001
- [5] Holland J H. Adaptation in nature and artificial system[M]. Ann Arbor: The University of Michigan Press, 1975
- [6] Tang K S, Man K F, Kwong S, et al. Genetic algorithms and their applications[J]. IEEE Signal Processing Magazine, 1996, 13(6): 22-37
- [7] Goldberg D E. Genetic Algorithms in Search, Optimization and Machine Learning[M]. Addison-Wesley, 1989
- [8] 杨华芬,魏廷.一种求解TSP问题的改进遗传算法[J].重庆科技学院学报:自然科学版, 2007, 21(5): 86-88