

基于遗传算法的神经网络油水层识别

曾黄麟¹, 李娟²

(1. 四川理工学院自动化与电子信息学院, 四川 自贡 643000; 2. 山西工商职业学院计算机系, 太原 030000)

摘要: 文章研究了基于遗传算法的神经网络油水层识别方法, 针对神经计算存在因输入信息空间维数较大而使网络结构复杂、训练时间长, 以及因冗余属性使网络拟合精度不高等缺点, 提出了基于粗集属性约简方法降低了输入信息的空间维数、减少了运算量和简化了神经网络的拓扑结构, 利用遗传算法提高神经网络的训练速度。实验结果表明: 将混合智能计算方法应用于油水层识别中效果显著, 其学习训练速度和拟合精度远优于传统 BP 神经网络算法。

关键词: 属性约简; 神经网络; 遗传算法; 油水层识别

中图分类号: TP18

文献标识码: A

引言

油水层识别是测井解释中的一项重要研究课题^[1-3]。由于用于油水层识别的测井信息和岩性、物性参数等属性诸多, 加上各个条件属性并不同等重要, 且常常存在着冗余信息, 因而在实际应用中, 必须进行属性约简以便于利用有用的信息来进行识别。

神经网络是一种模拟人类思维的智能方法^[4-6], 已在油水层识别中得到广泛应用, 它的优点在于具有大规模并行处理、信息分布存储等特点; 通过学习训练产生一个非线性映射, 自适应地对数据产生聚类; 具有较好的抑制噪声干扰的能力。其缺点主要是, 一般不能将输入信息空间的维数简化, 当输入信息空间维数较大时, 网络不仅结构复杂, 而且训练时间也很长。为此, 本文提出基于粗集属性约简和遗传算法改进的神经网络油水层识别方法, 具体创新工作为:

(1) 神经网络结构的简化。根据粗集属性约简算法, 删除冗余条件属性, 简化神经网络输入层节点数, 以达到简化神经网络结构简化的目的。

(2) 学习算法的优化。在 BP 神经网络训练中, 采用遗传算法训练 BP 神经网络的权值和阈值, 进而提高神经网络的学习训练速度, 克服 BP 神经网络由于初始权值和

阈值的随机选取而造成陷入局部极小等诸多问题。

1 油水层识别系统的约简

定义 1 油水层识别的知识系统 $I = (U, Q)$, 其中 U 是论域, Q 是属性集合, 分别为油水层识别的条件属性集 C 和识别的决策属性集 D , 且 $Q = C \cup D$, $C \cap D = \phi$ 。

定义 2^[7-8] $ind(P)$ 为由 $P(\subseteq Q)$ 决定的不可分辨关系。设 $p \in P$, 如果 $ind(P) = ind(P - \{p\})$, 则属性 p 是 P 中可约简的, 否则是不可约简的。由 P 中所有不可约简属性组成的集合称为 P 的核, 记为 $core(P)$ 。若 $R \subseteq P$, 有 $ind(R) = ind(P)$, 则 R 是 P 的一个约简, 记为 $red(P)$ 。显然, 所有约简集的交集等于 P 的核, 即 $core(P) = \bigcap red(P)$ 。

对于油水层样本集, 其属性约简步骤为:

(1) 连续属性离散化和决策属性泛化

油水层的决策属性是已知的, 即干层、水层、油水同层和油层类型, 将之依次分别泛化为连续非负整数。在选取完备的样本信息基础上, 按照决策属性的类型分别对条件属性进行相应的区间划分, 再利用中值、聚点原则求得分割点, 进而实现连续属性离散化。

(2) 求样本信息的核 $core(C)$

计算不可分辨关系 $ind(C)$ 、 $ind(C - \{c_i\})$ ($i = 1,$

2 ...), 求出可约简的属性, 进而得到 $core(C) = \cap red(C)$ 。

(3) 求相对最简约简集

由核 $core(C)$ 开始进行扩展, 在核中逐渐添加可约简的属性 c_i 得到集合 R , 直到 $ind(R) = ind(C)$, 则 R 为所求的相对最简约简集。

2 油水层识别学习算法的优化

油水层的识别可以通过训练学习一个神经网络来完成, 识别系统设计过程为:

(1) 样本信息预处理。选取完备的样本信息, 并进行标准化、归一化处理。

(2) 样本信息属性约简。先将连续属性离散化, 并将决策属性泛化, 再用粗糙集理论求出样本集的一个相对约简。

(3) 遗传学习优化网络。先依据样本的条件属性和决策属性来设计网络的输入和输出层节点数。再将属性约简后的样本集输入神经网络进行训练, 训练样本集的同时利用遗传算法优化神经网络的权值和阈值, 直到满足精度要求或规定的循环次数为止, 存储权值和阈值。

(4) 识别与输出。按照属性约简的结果, 剔除待识别信息中的冗余属性, 再经过预处理, 最后输入训练好的神经网络进行识别或分类, 并输出结果。

在 BP 神经网络训练中, BP 神经网络由于初始权值和阈值的随机选取而造成陷入局部极小等诸多问题。采用遗传算法训练 BP 神经网络的权值和阈值, 进而提高网络的学习训练速度。基于 GA 的训练算法简要描述:

(1) 初始化: 随机产生初始种群 $P = \{x_1, x_2, \dots, x_n\}$, 任一 $x_i \in P$ 为一神经网络, 它由一个权值向量和一个阈值向量组成。权值向量为 n 维实数向量, n 为所有的连接权的个数, 阈值向量也为 n 维实数向量 (不包括输入层神经元)。神经元编号采用自底向上、自左向右的方法 (包括输入层神经元)。其中 x_i 可用正态分布的小随机数来初始化。

(2) 评估: 根据随机产生的权值向量和阈值向量对应的神经网络, 对给定的输入集和输出集计算出每个神经网络的全局误差作为适应度, 即 $f = \sum_{p=1}^k \sum_{i=1}^n (d_{pi} - o_{pi})^2$ 。为了进行排序选择, 给已排好序的染色体设定选择概率, 概率确定采用 Michalewicz 提出的指数排序法, 即 $eval(\vartheta_i) = a(1-a)^{i-1}$, $a \in (0, 1)$, $i = 1, 2, \dots, pop_size$, $i = 1$ 意味着染色体是最好的, $i = pop_size$ 说明

是最差的。算法中取 $a = 0.05$ 。

(3) 排序并保留最好的染色体: 根据适应度从小到大排序, 即按染色体由好到坏, 染色体越好, 序号越小。并根据评估的结果给每个染色体按序号设定选择概率, 并保留最好的染色体。此最好染色体在后续的进化中可被更好的染色体所替代; 将新的适应度最小的染色体 (即最好的染色体) 和当前已保留的最好的染色体进行比较, 如果误差变小, 则保留新的最好的染色体。

(4) 选择过程: 选择过程是以旋转赌轮 pop_size 次为基础的, 每次旋转都为新的种群选择一个染色体, 赌轮为每个染色体按序号设定的概率选择染色体, 同时保证最佳个体的选择。这样设定选择概率, 概率并不是正比于染色体的适应度值, 可避免在较早代中一些超级染色体霸占选择过程, 而在较晚的代中种群集中在一起。算法表示为:

$$(i) \text{ 对每个染色体 } \vartheta_j \text{ 计算累积概率 } q_i: \begin{cases} q_0 = 0 \\ q_i = \sum_{j=1}^i eval(\vartheta_j), i = 1, 2, \dots, pop_size \end{cases}$$

(ii) 从区间 $(0, q_{pop_size})$ 中产生一个一致分布的随机数 r

(iii) 若 $q_{i-1} < r \leq q_i$, 则选择第 i 个染色体 ϑ_i ($1 \leq i \leq pop_size$);

(iv) 重复 b 和 c 操作共 pop_size 次, 这样可以得到 pop_size 个复制的染色体。如果在 $(pop_size - 1)$ 次选择过程中, 没有出现第一个最好的染色体, 则在第 pop_size 次选择时, 直接选择第一个最好的染色体。

(5) 交叉过程: 为确定交叉操作的父代, 从 $i = 1$ 到 pop_size 重复以下的过程: 从区间 $[0, 1]$ 中产生一个一致分布的随机数 r 如果 $r < p_c$, 则选择 ϑ_i 作为一个父代。如果用 $\vartheta'_1, \vartheta'_2, \vartheta'_3, \dots$ 表示上面选择的父代, 并把它们随机分成下面的对 $(\vartheta'_1, \vartheta'_2), (\vartheta'_3, \vartheta'_4), \dots$ 进行交叉。以 $(\vartheta'_1, \vartheta'_2)$ 为例对所进行的交叉操作解释如下: 从开区间 $(0, m)$ 中产生一个随机数 c (m 可根据不同的输出进行调节以便覆盖整个解集) 交叉采用如下形式: $x = c \times \vartheta'_1 + (2-c) \times \vartheta'_2, y = (2-c) \times \vartheta'_1 + c \times \vartheta'_2$ 其中 x 和 y 为后代, x 和 y 的替换对象分别为 2 个父代 ϑ'_1 和 ϑ'_2 。交叉完后, 为了考虑其优劣, 还原出对应的神经网络并进行性能评估。对同一对父代, 为了尽可能交叉出好的后代, 允许进行多次交叉。在交叉中只要有一个后代优于或两个后代都优于其对应的父代, 则替换其父代, 结束该对父代的交叉, 否则, 对该对父代进行下一次交叉。

(6)变异操作: 类似于交叉过程中父代的选择, 由 $i = 1$ 到 pop_size , 重复下列过程来选择变异的染色体: 从区间 $[0, 1]$ 中产生随机数 r , 如果, $r < p_m$, 则选择 q 作为一个父代。对选择好的每一个要变异的染色体, 为了尽可能进行好的变异, 允许进行多次变异。变异时, 首先随机生成一个与染色体的各权值、阈值同维数的向量 d_1 和 d_2 作为变异方向, 然后用父代染色体的权值、阈值对应的向量分别和向量 $m \times d_1$, $m \times d_2$ 相加 (仿真中取 $m = 2$)。对每次变异结果, 还原出神经网络并进行性能评估。如果后代优于父代则结束该父代的变异。否则, 对该父代进行下一次变异。

(7)结束条件: 如果网络误差满足要求, 或达到一定的进化代数, 停止进化, 输出进化结果; 否则, 转 (3) 步。

3 油水层识别系统实现

某一油田的 oilsk81井进行仿真试验, 试验数据来源于参考文献 [3], 进行油水层识别的条件属性为: 声波 (c_1)、中子 (c_2)、深测向电阻率 (c_3)、孔隙度 (c_4)、含油饱和度 (c_5) 和渗透率 (c_6), 即 $C = \{c_1, c_2, c_3, c_4, c_5, c_6\}$, 决策结果为: 干层、水层、油水同层和油层。令决策属性 $D = \{d\}$, $d = \{d_i = i \mid i = 0, 1, 2, 3\}$, 其中 0, 1, 2, 3 分别代表 4 个区间, 即 $[195, 210]$ 、 $[225, 230]$ 、 $[201, 240]$ 和 $[210, 225]$ 。

分别在各个区间内利用聚类的相关知识, 找到聚类中心最终得到离散区间的分割点。同理, 离散化其他 5 个条件属性, 将离散化后的样本信息进行预处理可得经离散化和预处理后的信息表。

按照前面的约简方法能够得到 $\{c_1, c_2, c_3, c_4\}$, $\{c_1, c_2, c_3, c_5\}$ 和 $\{c_1, c_2, c_3, c_6\}$ 这 3 个约简集, 其中 $\{c_1, c_2, c_3, c_4\}$ 为相对最简约简。

设前向神经网络的输出层有 4 个节点, 令干层、水层、油水同层和油层的期望输出分别为 $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, $(0, 0, 0, 1)$ 。

将表 1 中的属性 $\{c_1, c_2, c_3, c_4\}$ 作归一化处理:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

这里 x 是表 1 中属性的实际值, 即 $x \in [x_{\min}, x_{\max}]$, x_{\min} , x_{\max} 为最小, 最大值。

根据以上分析构建 4-15-4 的神经网络, 将归一化处理得到的条件属性值作为网络的输入, 将其决策属性作为网络分类识别的结果, 即 $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$ 和 $(0, 0, 0, 1)$ 。当满足遗传训练代数或达到网络的输出误差, 整个网络的学习训练便全部完成。

进行油水层识别, 识别误差如图 1 所示, 当训练代数达到第 100 代时, 识别误差达到预期精度。识别结果与传统的 BP 网络识别的结果见表 1。其中结论 0-3 分别表示干层、水层、油水同层和油层。

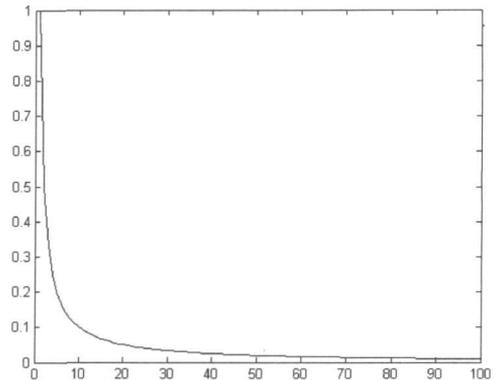


图 1 识别误差

表 1 对 oilsk85 井的识别结果比较分析

编号	试油结果	传统 BP	混合算法
1	0	0	0
2	0	0	0
3	0	0	0
4	0	2	0
5	0	0	0
6	0	0	0
7	1	3	1
8	1	1	1
9	2	1	1
10	2	2	2
11	2	0	2
12	2	2	2
13	3	3	3
14	3	3	3

由表 1 可得, 我们提出的混合智能学习方法的识别率高达 92.86%, 而传统 BP 网络的识别率只有 71.43%, 混合智能学习方法的应用效果明显优于传统 BP 网络。

5 结束语

本文采用了基于粗糙集的属性约简算法删除了用于油水层识别的冗余属性, 在神经网络的油水层识别系统中加入了遗传学习算法进行权值和阈值的训练, 从而克服了 BP 神经网络由于初始权值和阈值的随机选取而造成陷入局部极小问题, 利用 GA 在参数空间进行全局搜索, 取得较好的区域搜索与空间扩展的平衡。仿真结果显示混合智能学习算法用于油水层识别是切实可行、效果显著, 并且具有节省成本、提高处理速度等优点。

参考文献:

- [1] 陶果, 邹辉. 现代阵列声波测井数据处理和解释方法研究 [J]. 石油勘探与开发, 2000, 27(2): 76-78
- [2] 刘文龙. 相控声波测井仪器设计及资料处理方法 [J]. 石油仪器, 2006, 20(3): 14-17
- [3] 沈建国, 任月娥, 张宏敏. 高分辨率声波测井面临的问题及其对策 [J]. 石油地球物理勘探, 2006, 41(1): 111-115
- [4] Zeng Huanglin, Li Juan. An Oilwater Layer Recognition System Based on Composition Intelligence Computation [A]. Proceedings of the 2007 International Conference on Intelligent Systems and Knowledge Engineering [C]. 成都, 中国, 2007, 10, 989-992
- [5] Guo Haixiang, Zhu Kejun. Extracting fuzzy rules of oil-bearing formation based on soft computing [A]. (in Chinese), Proc of CRSSC-CW FCGIC' [C]. 2007, 239-254
- [6] Zeng Huanglin, Zeng Xiaohui. Studies on consistence of the knowledge simplification of an uncertain system, Proceeding of 8 International Conference on Cognitive Informatics [C]. HongKong 2009, 162-166.
- [7] Zeng Huanglin, Zeng Xiaohui. Reasoning Decision Rules of an Uncertain System [A]. The Fourth International Conference on Rough Sets and Knowledge Technology Proceeding of RSKT 2009 [C]. Australia Springer-Verlag Berlin Heidelberg 2009, 634-642
- [8] Zeng Huanglin, Huang Yan, Zeng Xiaohui. A New Approach of Attribute Reduction Based on Ant Colony Optimization [A]. Proceeding of Fifth International Conference on Natural Computation [C]. 2009, 33-37.

An Oilwater Layer Recognition Based on a Genetic Algorithm Neural Network

ZENG Huang-lin, LI Juan

(1. School of Automation and Electronic Information, Sichuan University of Science & Engineering, Zigong 643000, China
 2. Department of Computer Science, Shanxi College of Business and Technology, Shanxi 030000, China)

Abstract In this paper, a composition intelligence computing method is suggested for an oilwater layer recognition. The redundant condition attributes are reduced based on rough set attribute simplification algorithm, so that an oilwater layer neural network recognition system can be simplified to improve network composition. An optimization computation speed of neural network is improved by a BP learning with a genetic algorithm. Simulation result shows that the effect in oilwater layer recognition is improved by the composition intelligence computing method proposed here.

Key words attribute reduction; neural network; genetic algorithm; oilwater layer recognition