

文章编号: 0253-3782-(2002)06-0647-06

集成遗传算法及 BP 算法的 潜在震源区划分*

周庆叶洪

(中国北京 100029 中国地震局地质研究所)

摘要 以华南沿海地区为例, 集成遗传算法与 BP 算法进行潜在震源区的划分. 用遗传算法辅助人工神经网络的设计, 在无限的解空间中快速找到人工神经网络的最佳参数组合. 结果表明: 由该分类系统划分出的不同震级上限的潜在震源区分布, 反映了华南沿海地区地震环境与地震发生的内在规律性, 从而减少了人的主观判断影响.

关键词 遗传算法 BP 算法 潜在震源区

中图分类号: P315.5

文献标识码: A

引言

潜在震源区是指未来具有发生破坏性地震潜在可能性的地区. 潜在震源区的确定是概率法地震区划工作中的一个重要组成部分. 它是在地震活动性, 地震构造及地震区、带研究的基础上进行的. 潜在震源区划分方案的合理与否对地震区划最终结果有很大影响(国家地震局, 1996).

由于观察问题的角度、认识的差异、以及不同的研究者对于同一地区潜在震源区的划分往往不同, 当某一地区研究程度加深, 获得新的资料时, 潜在震源区需要重新划分或修改. 因此, 人工方法划分潜在震源区存在人的主观判断差异及工作效率不高问题.

现代高新技术的发展使潜在震源区的划分有可能突破传统模式, 人们可以利用计算机提供的强大数据处理能力, 在地理信息系统(GIS)、人工神经网络(ANN)和遗传算法(GA)等新技术的基础上, 实现动态划分潜在震源区的目的.

ANN 与 GA 同属人工智能技术, 均具有自学习、自组织、自适应及内在的并行性等智能特征. 人们已成功地将它们应用到那些难以用传统方法求解的复杂问题之中. 但作为 ANN 主要工具的反向传播算法(BP), 由于其本身的缺陷, 易陷于局部极小(余雪丽等, 1996); 而 GA 由于编码方案的限制, 由其控制的搜索可接近最佳参数组合在参数空间的位置, 但难以确定它们的确切位置(陈国良等, 1996). 将 GA 与 BP 算法结合、优势互补, 用遗传算法辅助神经网络的设计已被认为是最有前途的一种方法(Whitly, 1995; 潘正君等, 1998; Leung, 1997).

* 中国地震局“十五”重点项目(地震安全性评价研究与结构抗震研究)和科技部专项(2001DEB30078)共同资助.
2001-12-17 收到初稿, 2002-08-05 收到最后修改稿, 2002-10-07 决定采用.

1 GA 与 ANN 的集成

1.1 利用 GA 与 BP 的组合辅助设计 ANN

描述一个 ANN 模型的参数主要有：网络层数、各层神经元数、神经元的互连方式、连接权值以及传递函数等。设计的 ANN，实质上就是根据给定的性能评价标准搜索确定适合于解决某个问题或某类问题的网络参数的组合(马玉书，1998)。

GA 与 BP 可以有多种不同的组合方法，本文采用先 GA 后 BP 的组合方式。先用 GA 反复优化描述 ANN 模型的参数组合，直到适应度函数的平均值不再有意义地增加或内符合率达到指定最大值为止。此时，解码串得到的参数组合已经充分接近符合特定应用需要的最佳参数组合。在此基础上，再用 BP 算法利用经过精选的训练集对网络参数进行优化。

1.2 人工神经网络连接权的进化

1.2.1 GA 优化 ANN 的过程

神经网络连接权的整体分布包含着神经网络系统的全部知识，传统的权值获取方法是先赋予确定的初始权值或由计算机随机给出，在训练中逐步调整，最终得到一个较好的权值分布。由于算法的缺陷，训练时间有可能过长或因陷入局部极小值而得不到适当的权值分布。

用遗传算法优化神经网络连接权的过程如下：

1) 随机产生一组分布，采用某种编码方案对该组中的每个权值(或阈值)进行编码，进而构造出一个个码链(每个码链代表网络的一种权值分布)。在网络结构和学习规则已定的前提下，该码链就对应一个权值和阈值取特定值的神经网络。

2) 对所产生的神经网络计算它的误差函数，从而确定其适应度函数值，误差越大，则适应度越小。

3) 选择若干适应度函数值最大的个体，直接遗传给下一代。

4) 利用交叉和变异等遗传操作算子对当前一代群体进行处理，产生下一代群体。

5) 重复 2)~4)，使初始确定的一组权值分布得到不断地进化，直到训练目标得到满足为止。

1.2.2 适应度函数的确定

在本项研究中适应度函数取 $F=1/e$ ，其中 e 为误差绝对值的和：

$$e = \sum_m \sum_k |Y_{mk} - \bar{Y}_{mk}| \tag{1}$$

式中， Y_{mk} 及 \bar{Y}_{mk} 分别为第 m 个训练样本的第 k 个输出节点的期望输出与实际输出， e 的计算过程即为神经网络的正向处理过程。

1.2.3 GA 的内在并行性

由于GA采用种群的方式组织搜索，因此，它可以同时搜索解空间内的多个区域。通过遗传操作，实质上GA可以进行比种群规模 N (本文中 $N=100$)大得多的大约 $O(N^3)$ 次有效搜索(潘正君等，1998)。

2 基于 GIS 与人工智能的华南沿海潜在震源区划分

2.1 数据准备

潜在震源区的划分通常遵循“地震重复”及“构造类比”原则。“地震重复”意指过去发生过强震的地区将来仍有可能发生同样强烈的地震；“构造类比”则指的是地质构造条件相似的地区，有可能发生震级相近的地震(国家地震局，1996)。这两条原则在用 ANN 划分潜在震源区时体现在 ANN 训练集的选择及网络判别中。

本文采用的 BP 算法是有监督的学习算法，其过程是先提取曾经记录到相应震级地震的地区各网格单元的特征信息，经筛选后作为学习样本集，再将经过训练好的网络用于判别其它地区的网格单元。这样，与训练样本集中地质构造背景类似的网格单元就可能被划分到同一类别，从而达到划分潜在震源区的目的。

2.1.1 ANN 特征指标的选取

特征指标的选择对于网络能否成功运行至关重要，网络输入模式的节点个数，及各节点所代表的特征指标需重点把握。

基于华南沿海地区地震地质、地震活动性以及以往地震区划的研究成果，考虑现阶段该地区地震环境研究的水平，在本项研究中提取 8 个特征指标：① 是否有主要活动断裂通过；② 是否处在主要断裂与北西向断裂交汇部位；③ 是否有北东向断裂通过；④ 是否有北西向断裂通过；⑤ 是否有活动断裂相交；⑥ 是否是第四纪活动断裂；⑦ 是否与新生代盆地相交；⑧ 30 km 范围内是否有 6.5 级以上地震发生。

主要活动断裂在此指的是北东向广东滨海断裂带及南海北部陆架南缘断裂带。

2.1.2 建立人工神经网络

针对 3 个震级档： $M_0 \geq 7.0$ ， $6.0 \leq M_0 < 7.0$ 和 $5.0 \leq M_0 < 6.0$ ，分别用 3 个各自独立的 3 层人工神经网络建立地震环境与各单元震级上限的关系。ANN 采用 $8 \times 4 \times 1$ 的结构，即有 8 个输入节点、1 个输出节点和 4 个隐节点。

具体采取的步骤是依 $M_0 \geq 7.0$ ， $6.0 \leq M_0 < 7.0$ 和 $5.0 \leq M_0 < 6.0$ 的次序分别训练样本集，被判别为高震级档的单元将不参加下一档的判别；对于一个 ANN，先用 GA 算法粗略估算神经元间的连接权及阈值，在此基础上，再用 BP 算法精确确定 ANN 的各种参数，使网络训练集的内符准确率达到 100%。

用 GA 优化 ANN 时，适应度函数取 $F=1/e$ 。随机产生 100 个群体，交叉概率在 0.1~0.9 之间选择，而变异概率可以在 0.001~0.01 间取值，演化 10 代后选择内符准确率最高的一个群体，译码后的权值及阈值分布就是网络进一步优化的基础。内符准确率以期望输出与实际输出的误差绝对值 $s-error < 0.3$ 为标准。

利用 BP 算法进行网络的再训练，网络初始权值与阈值采用 GA 算法后的处理值，而不是随机产生的，这样就避免了网络陷入局部极小值或不收敛的情况发生。

2.2 ANN 的建立

以模式不重复为原则，经 GIS 处理筛选得到 $M_0 \geq 7.0$ ANN， $6 \leq M_0 < 7.0$ ANN 和 $5.0 \leq M_0 < 6.0$ ANN 三个神经网络的训练样本集，分别包括 21，57，15 个单元，样本集中包括已发

生相应震级档地震的所有单元模式.

2.2.1 遗传算法优化 ANN

对 3 个独立的网络分别依次进行遗传算法优化, 最终的技术指标如表 1 所示. 同时得到 3 组网络权值与阈值分布, 即 3 个神经网络的初始优化组合(周庆, 2001).

表 1 用遗传算法优化 ANN 的主要技术指标

ANN	网络迭代 次数	交叉 概率	变异 概率	交叉操作 次数	变异操作 次数	内符 准确率
$M_0 \geq 7.0$	29	0.1	0.004	60	0	95.24%
$6.0 \leq M_0 < 7.0$	10	0.9	0.006	2 422	13 993	78.95%
$5.0 \leq M_0 < 6.0$	35	0.7	0.001	1 120	11 406	80.00%

2.2.2 BP 算法训练 ANN

利用 BP 算法进行网络的再训练, 网络初始权值与阈值采用上节 GA 所提供的组合, 网络训练内符准确率均达 100%, 最终得到进一步优化了的网络参数组合, 从而建立 3 个可用于判别的 ANN.

2.2.3 ANN 测试与判别

为检验 ANN 的稳定性, 在区域内分别提取 3 个不同级别的测试集, 测试样本集中的样本可以重复, 因此样本数可以很多.

$M_0 \geq 7.0$ ANN、 $6.0 \leq M_0 < 7.0$ ANN和 $5 \leq M_0 < 6.0$ ANN三个测试集分别包括 36, 216, 218 个单元. 测试结果表明, 网络比较稳定, 误判率分别为 5.6%、12.5%和 18.4%.

将经过训练、测试后的网络应用于全区判别, 从 5 579 个单元中得到 142 个 $M_0 \geq 7.0$ 单元, 从 3 123 个单元中得到 879 个 $6.0 \leq M_0 < 7.0$ 单元, 从 2 244 个单元中得到 1 291 个 $5.0 \leq M_0 < 6.0$ 单元.

3 讨论和结论

由本系统划分出的不同震级档次网格单元, 经平滑处理后得到图 1 所示的潜在震源区分布图. 将断层分布图叠加其上, 可以看出: 由本方法划分出的 7 级以上潜在震源区, 基本上分布在北东东向广东滨海断裂带及南海北部陆架南缘断裂带与北西向断裂相交汇的部位; $6.0 \leq M_0 < 7.0$ 潜在震源区则沿上述两个断裂带除 7 级潜在震源区之外的地区、断裂相交的部位以及第四纪断陷盆地; 而 $5.0 \leq M_0 < 6.0$ 潜在震源区则基本上沿北东或北东东向断裂分布. 由此可知, 潜在震源区的分布主要由活动构造控制, 可能的最大震级依构造部位不同而发生变化.

本项研究所用的方法较好地反映了华南沿海地区地震环境与地震发生的内在规律性, 基本上符合以往人们对该地区中强地震的发生条件所做的结论. 例如, 在研究区内发生 $M \geq 7$ 地震的构造部位是: ① 区内两条北东东向的主释能带: 北东东向广东滨海断裂带及北东东向南海北部陆架南缘断裂带; ② 在上述两条北东东向断裂带上与区域性北西向断裂带交汇的部位, 或被区域性北东向断裂带截切的地方.

系统采用的是一种自学习、自适应方法. 根据已发生中强地震地区的地震地质条件, “学习”中强地震发生与活动构造背景之间的内在联系(这种联系被记录在神经网络处理单

元间的连接权值与阈值上)，从而建立一种判别模型，判断未发生过相应震级以上地震的地区发生该档震级地震的可能性。

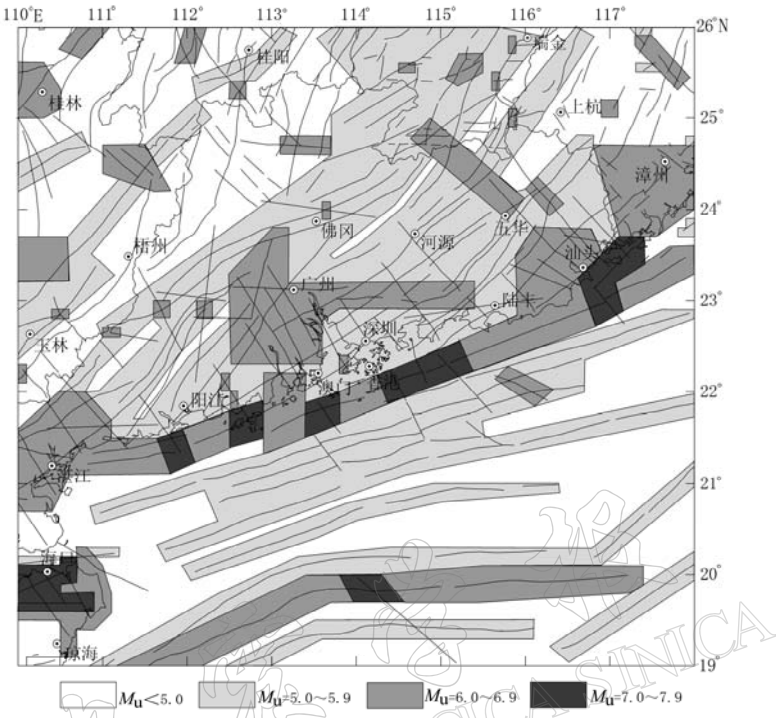


图 1 研究区内潜在震源区与断裂分布图

本研究的实践证明，用遗传算法辅助人工神经网络的设计，在无限的解空间中迅速找到人工神经网络的最佳参数组合。两种人工智能技术互为补充，可以有效地解决难以用传统方法进行的复杂分类问题。该方法在地震预报、地球物理等其它领域有望得到进一步的应用。

以上是笔者将遗传算法及 BP 算法两种人工智能技术应用于工程地震领域的一次初步尝试。还有一些技术性问题可以在下一步工作中完成，如：网络结构的简化，观察隐层单元个数减少时，网络的运行情况；试验能否单独使用 GA 解决问题等等。

参 考 文 献

陈国良，庄镇泉，王熙法，等。1996. 遗传算法及其应用[M]. 北京：人民邮电出版社，1~25，241~258
国家地震局。1996. 中国地震烈度区划图(1990)概论[M]. 北京：地震出版社，89~100
马玉书。1998. 人工智能及其应用[M]. 北京：石油大学出版社，424~480
潘正君，康立山，陈毓屏。1998. 演化计算[M]. 北京：清华大学出版社，124~131
余雪丽，孙承意，冯秀芳，等。1996. 神经网络与实例学习[M]. 北京：中国铁道出版社，145~174
周庆。2001. 人工智能技术在确定潜在震源区中的应用——以华南沿海地区为例[D]：[学位论文]. 北京：中国地震局地质研究所，73~90
Leung Y. 1997. Intelligent Spatial Decision Support System[M]. Berlin: Springer-Verlag, 229~268
Whitly D. 1995. Genetic Algorithms and Neural Networks[M]. Canada: John Wiley & Sons, 203~216

DEMARCATIION OF POTENTIAL SEISMIC SOURCES ON INTEGRATION OF GENETIC ALGORITHM AND BP ALGORITHM

Zhou Qing Ye Hong

(Institute of Geology, China Seismological Bureau, Beijing 100029, China)

Abstract: In this paper potential seismic sources in coastal region of South China are identified by integration of genetic algorithm and back propagation (BP) algorithm. GA is used for finding the best parameter combination rapidly in an infinite solution space for artificial neural networks (ANN). The results show that the distribution of potential seismic sources with different upper magnitude demarcated by this classifier is mostly satisfied the intrinsic relationship between seismic environment and earthquake occurrence, with less effect from subjective judgment of human being.

Key Words: genetic algorithm; BP algorithm; potential seismic sources

地 震 学 报
ACTA SEISMOLOGICA SINICA