**基于熵权法和神经网络的信息系统安全性评价**

**摘要：**

为了提高信息系统安全性评价的准确性，提出了基于熵权法和神经网络相融合的信息系统安全性评价模型（熵权-BP），避免了主观方法确定评价指标权重的随意性，弥补了传统评价方法无法处理非线性关系的不足。验结果表明，熵权-BP模型不仅加快了BP神经网络收敛速度，而且还提升了信息系统安全性评价精度，能够更好地适应当前信息系统的复杂性。

关键词：信息系统 熵权法 神经网络 评价模型 主观方法 非线性

**0 引言**

随着各行业对信息系统的依赖度日益增加，当信息系统发生安全事故时，会给企业及国家安全带来灾难性的后果。比如，浙江省温州有线电视网络系统遭黑客攻击，对群众生活造成了不良影响；乌克兰电力信息系统遭受到恶意代码攻击, 导致十万用户大停电。因此，信息系统的安全状况成为当前研究的热点问题，如何准确地评价信息系统安全性具有重要的研究价值和现实意义。

信息系统安全性评价[1]主要分为定性分析和定量分析两种方式，由于影响信息系统的安全因素难以量化，大多数学者采用层次分析法、模糊理论、故障树等[2,5]定性与定量相结合的方法建立信息系统安全性评价模型。尽管这些传统的方法有较深的理论基础，但计算复杂、评价结果受主观因素影响较大。近年来，D-S证据理论[6]、支持向量机[7]、神经网络[8]、贝叶斯网络[9]等人工智能算法已被广泛应用在信息系统安全性评价领域。比如马丽仪[10]采用模糊神经网络评价信息系统的安全风险，其模糊处理阶段的隶属度矩阵和权重均采用专家评判，客观性较差，在一定程度上影响整个模型的实用性；赵保华[11]采用层次分析法和神经网络构建信息系统安全评价模型，层次分析法确定评价指标重要性时完全依赖评价者的主观判断，可能会误剔重要指标，而且评价指标较多容易出现残缺判断矩阵。

熵权法是客观确定权重的一种方法，利用实际数据通过计算得到权重，避免了主观方法确定指标权重的严重缺陷，使指标权重更具有科学性和说服力。相比其他人工智能算法，BP神经网络（Back-Propagation Neural Network）具有较强的自学习、自适应能力，可以实现输入到输出的高度非线性映射关系，训练好的BP模型能解决具有类似特点的问题。BP神经网络可以有效的解决复杂信息系统各评价指标之间的互相影响问题，能够克服传统的评价方法只能处理评价指标到评价结果的线性关系。为了进一步提高评价的准确性，提出熵权法和BP神经网络相结合的信息系统安全性评价方法（熵权-BP），实验结果表明,熵权-BP神经网络评价方法提高了信息系统安全的评价精确度，使评价结果更接近信息系统安全的真实情况。

1 BP神经网络模型

BP神经网络即误差反向传播神经网络，目前被广泛用在综合评价领域。BP神经网络可以处理“多输入-多输出”的高度非线性的映射关系。BP神经网络通过输入信息号正向计算误差和反向调整网络权值和阈值这两个阶段不断的交替进行，神经网络就会获得不断的学习机会。当误差信号减到预定值或学习次数达到最大规定次数，学习过程结束，此时BP神经网络内部已经存储了大量的知识。典型的BP神经网络包括输入层，隐含层和输出层如图1所示：



**图1 BP神经网络信息系统安全性评价结构**

将熵权法和BP神经网络的优点相结合，构建信息系统安全性评价模型。熵权法计算得到评价指标的权重，选择权重较大的评价指标作为BP神经网络的输入，可以降低BP神经网络的输入维数，提高输入数据的质量，使得BP神经网络在学习时具有较高的识别能力，从而使评价结果更接近于实际值。

**2. 熵权-BP神经网络信息系统安全性评价过程**

**步骤1**根据《信息安全技术 信息系统安全等级保护测评要求》，科学合理的构建信息系统安全性评价体系；

**步骤2** 收集已完成评价的信息系统样本数据，利用熵权法计算评价指标的权重；

**步骤3** 通过对评价指标权重的排序，筛选出权重相对较大的评价指标；

**步骤4**将筛选出的重要评价指标作为BP神经网络的输入，即为BP神经网络输入层神经元的个数；

**步骤5** 对评价指标数据进行归一化处理，方便BP神经网络模型的训练；

**步骤6** 通过BPNN对收集的信息系统安全评价样本数据进行训练，得到稳定的信息系统安全评价模型；

**步骤7** 用测试样本对熵权-BP神经网络信息系统安全评价模型进行检验**，**并分析其性能。

熵权-BP神经网络信息系统评价流程如2图所示：

**图2 熵权-BP神经网络信息系统评价过程**

**3熵权-BP神经网络信息系统安全性评价模型的设计**

**3.1构建评价指标体系**

评价信息系统的安全性首先是构建评价指标，即信息系统安全影响因素。影响信息系统安全的因素复杂、多样化，使得提取评价指标十分困难，而评价指标的选取是否科学合理，直接影响评价结果的真实性。评价指标的选取是否科学合理，直接影响评价结果的真实性。目前，信息系统安全评价主要是基于《信息安全技术 信息系统安全等级保护测评要求》，从技术和管理两大层面对信息系统的安全问题进行全方位的评价。信息系统安全性评价指标体系的递阶层次结构如图3所示。

**图3 信息系统安全性评价指标体系**

**3.2熵权法确定评价指标权重**

在信息系统安全性评价时，各评价指标在整个评价体系中的重要程度不同，指标权重的设定是否合理直接会影响评价结果的准确性。确定权重的方法主要有德尔菲法、层次分析法和熵权法等，其中德尔菲法、层次分析法属于主观赋权法，主观随意性较大，而熵权法属于客观赋权法，利用实际数据计算得到权重，使评价指标权重更具有科学性和说服力。由于不同信息系统的评判专家可能不同，受专家主观因素影响，同一指标在不同信息系统的权重可能也不同。因此，本文根据各评价指标在不同信息系统的变异度，采用熵权法计算出各评价指标的权重，避免了人为确定评价指标权重的严重缺陷，而得到较为客观的指标权重。

**3.2.1构建评价指标矩阵**

从已评价的信息系统中筛选个具有代表性的信息系统，与个评价指标形成原始数据矩阵：

（1）

其中为第个信息系统的第个指标的实际评价结果值。

**3.2.2 计算评价指标的权重**

（1）根据2.2.1节构建的评价指标原始数据矩阵，计算第个评价指标下第个信息系统的评价指标值的比重。

（2）

（2）计算第个评价指标的熵值

 （3）

由熵的极值性可知，若某评价指标的熵越小，说明该评价指标在不同信息系统中的变异程度越大 ，提供的信息量越多，在综合评价中起的作用就越大，其权重也应该越大。

（3）用对各评价指标的熵值进行归一化处理：

 （4）

（4）引入差异系数 来正向度量评价指标的熵值，得到评价指标的权重：

（5）

式中，。

**3.2.3 筛选评价指标**

通过以上步骤得到所有评价指标权重并对其进行排序，剔除在信息系统安全评价中几乎不起作用的评价指标，筛选出权重相对较大的评价指标作为BP神经网络的输入，减少BP神经网络的输入维数，提高输入数据质量，较快BP神经网络收敛速度。

**3.3 BP神经网络学习算法设计**

（1）给定组样本的输入和输出分别为、；

（2）BP神经网络隐含层和输出层正向传递；

隐含层样本的输入和输出分别为：

 （6）

其中表示隐含层神经元的偏置值。

输出层样本的输入和输出分别为：

 （7）

其中表示输出层神经元的偏置值。

（3）使用联合算法，反向修正BP网络权值和偏置值：

（8）

（9）

（10）

式中，为对求导，为学习速率，加快网络的收敛速度；为动量因子，避免BP网络陷入局部最优。

（4） 组样本拟合误差函数：

（11）

式中为期望输出值，为网络实际输出值，为误差精度，是很大的整数，即网络训练的最大次数。当网络误差达到精度要求或者达到循环要求的最大次数，BP网络训练结束，否则转向步骤(2)，继续训练直到满足条件为止。

**4 实例研究**

**4.1 数据来源**

为检验EWM-BP神经网络信息系统安全性评价模型的性能，对其进行实验并分析结果。收集民航已完成安全性评价的30个信息系统作为样本数据,对每个评价指标数据用最大最小法归一化处理得到新数据样本，如表1。将30个样本分为22个训练集和8个测试集两部分。

表1 信息系统安全评价数据

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 样本编号 |  |  |  |  |  |  |  |  |  |
| 1 | 68.4 | 59.43 | 68.43 | 62.7 | 85.08 | 74.68 |  | 80.97 | 0.75 |
| 2 | 75.4 | 65.2 | 75.67 | 82.84 | 84.48 | 73.32 |  | 84.62 | 0.81 |
| 3 | 61.64 | 62.59 | 75.11 | 65.8 | 79.68 | 94.68 |  | 87.37 | 0.67 |
| 4 | 69.28 | 59.66 | 72.35 | 60.27 | 86.08 | 80.68 |  | 76.87 | 0.75 |
| 5 | 70.16 | 73.53 | 75.36 | 72.66 | 77.32 | 81.6 |  | 68.25 | 0.71 |
| 6 | 91.93 | 67.29 | 63.4 | 82.24 | 91.7 | 100 |  | 78.72 | 0.83 |
|  |  |  |  |  |  |  |  |  |  |
| 30 | 78.74 | 75 | 61.6 | 62.17 | 81.16 | 94.68 |  | 84.85 | 0.76 |

**4.2 熵权法计算权重**

（1）使用已完成安全评价的30组评价指标数据形成一个原始数据矩阵,即



（2）通过3.2节的计算步骤，得出各评价指标的权重：

按评价指标重要度排序后：



权重为0.005以下的评价指标值的改变对整个评价结果几乎没有影响，故将其剔除，剩下8个权重较大评价指标作为BP神经网络模型的输入。

**4.3 EWM-BP神经网络建模**

EWM-BP神经网络模型共有三层，输入层神经元的数量为通过熵权法筛选出较大权重的评价指标个数，输出层只设定一个神经元，代表信息系统安全性评价结果。确定隐含层神经元的个数是整个建模成功的关键，借助经验公式：

（12）

式中：表示输入层神经元个数，的常数。

隐含层和输出层神经元的输入和输出之间的非线性关系的确定采用Sigmoid函数。设定样本训练的学习率为0.5，动量因子为0.3，最大学习次数为6000，拟合误差为。

**4.4 EWM-BP模型训练**

将权重较大的评价指标数据采用最大最小法归一化处理后输入到BP模型中进行学习训练，当训练4000次时，拟合误差为0.000237，此时EWM-BP模型已达到拟合误差要求。为了比较EWM-BP模型与BP模型的学习能力，同时对两个模型进行训练，其收敛对比曲线如图4所示。

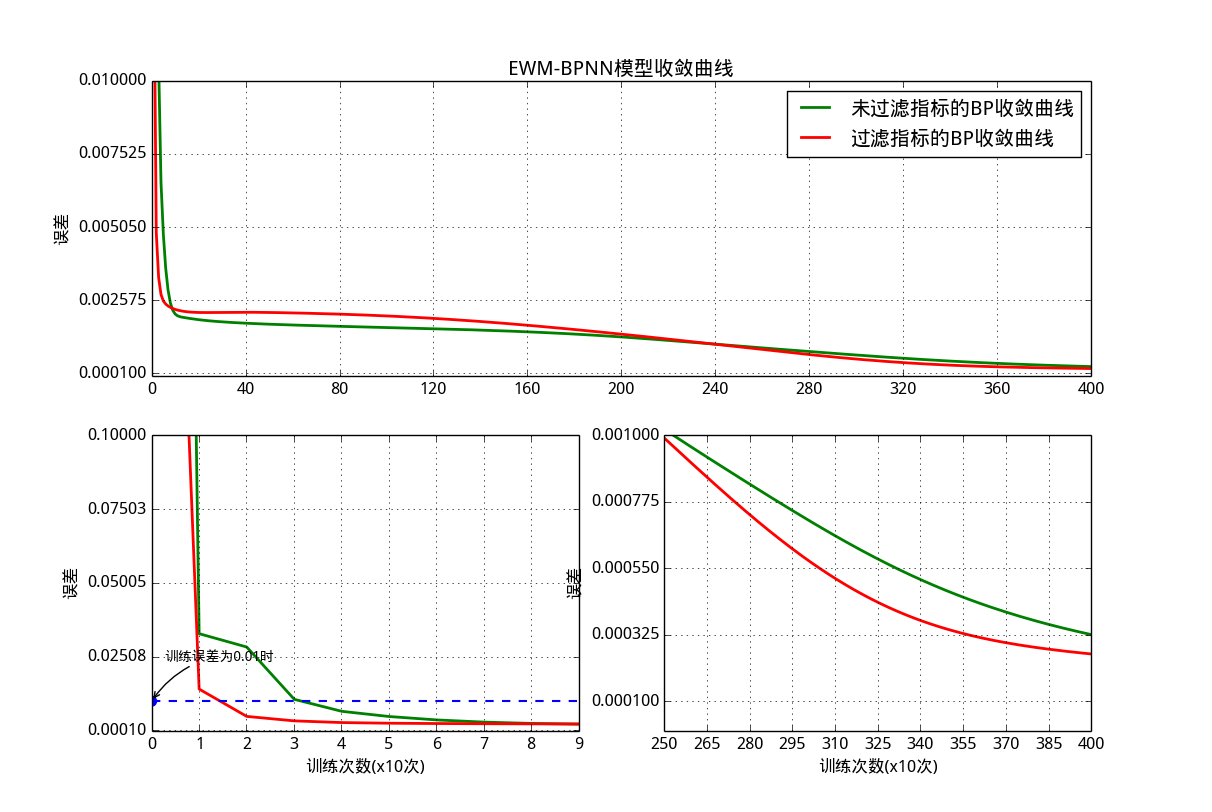


图4 BP神经网络训练集的收敛过程

由图4可知，EWM-BP模型在训练初期的收敛效果明显比单纯使用BP神经网络的效果好。当训练次数达到30次时，误差已经接近0.001，而BP需要迭代60次才能达到相同的误差，同时，当训练次数从2500-4000次时，EWM-BPNN的收敛速度也明显比BP神经网络快，最终训练完成后EWM-BPNN得到的误差为0.000237，而BPNN的误差为0.000325，精度提高了37.13%。

训练结果表明，EWM-BP模型通过对筛选出8个权重较大的评价指标样本数据进行训练，可以有效的减少迭代次数，并快速收敛到设定的误差，既可以提高BP神经网络的训练速度，又能保留重要指标的特征，减少非重要指标对BP神经网络模型的干扰，提高BP神经网络评价模型的准确度。

**4.5 EWM-BP模型检验**

将测试样本分别输入到BPNN、AHP-BP和EWM-BP模型中得到信息系统安全的网络实际输出值与期望值的对比关系如图5所示，同时表2给出了各模型的信息系统安全性评价精度，可以得到以下结论:

（1）使用EWM-BP和AHP-BP得到的信息系统安全评价结果比直接使用BPNN模型得到的评价结果更接近期望值，这是因为以上两种算法都对信息系统的指标进行筛选，过滤掉对评价结果没有重要影响的评价指标，使得BP神经网络在反向传播过程中更加逼近真实权值，从而使最终评价结果更接近期望值。

（2）EWM-BP模型的评价结果精度高于AHP-BP模型的评价结果，这是因为EWM-BP模型通过样本数据计算得到指标的权重，更加客观、真实的反映评价指标的权重，而AHP-BP模型通过专家构造判断矩阵计算指标权重并筛选出重要指标，主观因素较强，同时AHP算法可能存在判断矩阵的不一致性问题，使得评价指标的筛选不稳定，模型效果受主观因素影响较大。

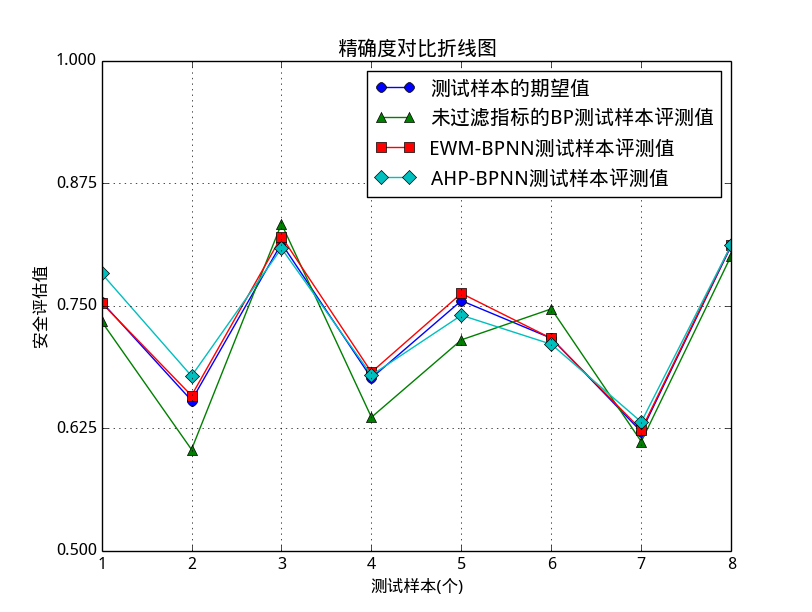


图5 各模型的网络输出值和期望值之间的对比

|  |  |  |  |
| --- | --- | --- | --- |
| 实际值 | BPNN | AHP-BP | EWM-BP |
| 0.75 | 0.69 | 0.78 | 0.74 |
| 0.65 | 0.58 | 0.60 | 0.63 |
| 0.81 | 0.75 | 0.78 | 0.80 |
| 0.67 | 0.58 | 0.61 | 0.68 |
| 0.75 | 0.71 | 0.73 | 0.74 |
| 0.71 | 0.65 | 0.71 | 0.71 |
| 0.62 | 0.55 | 0.63 | 0.62 |
| 0.89 | 0.80 | 0.85 | 0.88 |

表2 各模型的信息系统安全评价精度对比

|  |  |
| --- | --- |
| 模型 | 评价精度/% |
| BPNN | 91.67 |
| AHP-BP | 93.25 |
| EWM-BP | 96.14 |

**5 结束语**

本文将熵权法和神经网络相融合，构建了一个针对复杂信息系统的安全性评价模型。实验结果表明，EWM-BP模型通过采用熵权法计算各评价指标的权重，筛选出重要指标，提高了评价指标的质量，避免了主观因素的影响，同时结合具有高度非线性处理能力的BP神经网络，解决了各评价指标之间的相互影响，弥补了传统评价方法无法处理非线性关系的不足，提高了评价的精度，能够更好地适应信息系统的复杂性。但此方法还存在不足，采集样本的数量受到限制，样本的正确性不好界定，需进一步的研究改进。

参考文献：

[1] 孙强. 信息安全风险评估模型的定性与定量对比研究[J]. 微电子学与计算机, 2010, 27(6): 92-96.

[2] 赵刚, 刘换. 基于多层次模糊综合评判及熵权理论的实用风险评估[J]. 清华大学学报: 自然科学版, 2012, 52(10): 1382-1387.

[3] 顾兆军, 韩迎亚. 可量化的信息系统安全性水平评估[J]. 计算机工程与设计, 2016, 37(7): 1729-1733.

[4] 林洋. 信息系统安全风险评估模型及其应用研究[D]. 大连: 大连海事大学, 2013.

[5] Rosqvist T, Molarius R, Virta H, et al. Event tree analysis for flood protection—An exploratory study in Finland[J]. Reliability Engineering & System Safety, 2013, 112: 1-7.

[6] 王姣, 范科峰, 莫玮. 基于模糊集和DS证据理论的信息安全风险评估方法[J].计算机应用研究，2017，34(11)

[7] 伍浏阳. 因子分析和支持向量机的信息系统风险评价[J]. 微电子学与计算机, 2016, 33(2): 144-148.

[8]高妮, 高岭, 贺毅岳, 等. 基于贝叶斯攻击图的动态安全风险评估模型[J]. 四川大学学报: 工程科学版, 2016, 48(1): 111-118.

[9] 崔丹丹. 模糊神经网络在信息系统安全评价中的应用[J]. 科学技术与工程, 2014 (7): 212-215.

[10] 马丽仪, 张露凡, 杨宜, 等. 基于模糊神经网络方法的信息系统安全风险评价研究[J]. 中国安全科学学报, 2012, 22(5): 164-169.

[11] 赵保华. 层次分析法和神经网络的信息系统风险评估[J]. 微电子学与计算机, 2015, 32(10): 163-166.

[12] 张俊林. 网络安全风险动态评估方法研究[J]. 计算机仿真, 2016 (10): 356-360.

[13] 张利, 彭建芬, 杜宇鸽, 等. 信息安全风险评估的综合评估方法综述[J]. 清华大学学报: 自然科学版, 2012, 52(10): 1364-1369.

[14] 马刚, 杜宇鸽, 荣江, 等. 基于威胁传播的复杂信息系统安全风险评估[J]. 清华大学学报 (自然科学版), 2015, 54(1): 35-43.

[15] 窦志武, 李红巍, 熊琦. 基于熵权法和神经网络的口岸物流综合能力评价方法研究[J]. 计算机科学, 2015, 2.