·信息安全技术·

**基于熵权法和神经网络的信息系统安全性评价**

顾兆军1，辛倩1

（1.中国民航大学，信息安全测评中心，天津 300300）

**摘 要：**在当前复杂的信息系统安全性评价过程中，存在着很多不确定的因素，针对主观方法确定指标权重的随意性和不确定性，以及传统方法无法处理评价指标到评价结果非线性关系，提出了一种将熵权法和神经网络相结合的信息系统安全性评价模型（熵权-BP）。实验结果表明，与其他同类型评价模型相比，该模型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

**中图法分类号:** TP398.1 **文献标识码:**A **文章编号:** （由本刊工作人员填写）

Information System Security Evaluation Based on Entropy Method and Neural Network,

Zhao Jun Gu1, Qian Xin1

（1.Civil Aviation University of China, Information Security Evaluation Center, Tianjin 300300）

**Abstract:** There are many uncertain factors in the current complex information system security evaluation process . For the arbitrariness and uncertainty of the weight of the evaluation index of subjective determination and the traditional method can not deal with the non-linear relationship between the evaluation index and the evaluation result, a security evaluation model (entropy weight - BP) is proposed to combine the entropy weight method and the neural network. Compared with other evaluation models of the same type the experimental results show that, the model can improve the accuracy of information system security evaluation and can better adapt to the complexity of the current information system.

**Key words:** information system; entropy weight method; neural network evaluation model; subjective method; nonlinearity

# 引言

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[2,4]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，国内外学者采用DS证据理论[5]、支持向量机[6]、神经网络[7]、贝叶斯网络[8]等人工智能算法构建信息系统安全性评价模型，取得了较好的评估结果。张利[9]等人，提出了4种改进的具有代表性的信息安全风险评估方法,包括传统数学方法和人工智能方法，并对这4种方法的优缺点进行了对比分析；马丽仪[10]等人采用模糊神经网络评价信息系统的安全风险，将风险因子进行模糊化后输入到BP神经网络训练，提高了BP神经网络对模糊性的识别能力，但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。赵刚[12]等人使用熵权法计算指标权重，使模糊合成中的权重更加客观，但是当信息系统有较多不确定因素时，缺乏学习和自适应能力。

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，本文提出熵权法和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 熵权-BP信息系统安全性评价过程

利用已完成安全性评价的信息系统作为样本数据构建评价指标矩阵，采用熵权法计算得到各评价指标权重，将权重较大的评价指标作为BP神经网络的输入，通过自适应能力与非线性映射能力强的BP神经网络训练学习得到最优的信息系统安全性评价模型。最后利用测试样本检验熵权-BP模型性能，得到信息系统安全性评价结果。熵权-BP模型评价过程如1图所示：



图1 熵权-BP模型评价过程

# 2 熵权-BP神经网络信息系统安全性评价模型的设计

## 2.1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。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。因此，分析当前信息系统安全的研究现状，结合专家意见，从技术和管理两大层面构建了递阶层次结构的信息系统安全性评价指标体系，如图2所示。



图2 信息系统安全性评价指标体系

## 2.2 熵权法筛确定指标权重

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。

2.2.1构建评价指标矩阵

从已评价的信息系统中筛选m个具有代表性的信息系统，与n个评价指标形成原始数据矩阵R=(rij)nm：

（1）

其中rij(i=1,2,..,n;j=1,2,…,m)为第j个信息系统的第i个指标的实际评价结果值。

2.2.2 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R，计算第i个评价指标下第j个信息系统的评价指标值的比重pij。

（2）

2）计算第i个评价指标的熵值ei

 （3）

由熵的极值性可知，若某评价指标i的熵ei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ln(m)对各评价指标的熵值进行归一化处理：

 （4）

4）引入差异系数1-ei 来正向度量评价指标i的熵值，得到评价指标i的权重：

（5）

式中，0≤wi≤1。

## 2.3 熵权-BP神经网络的构建

BP神经网是一个“多输入-多输出”的高度非线性的映射关系，能够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。为了避免BP神经网络在学习过程中容易过拟合，采用熵权法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，从而提高BP神经网络学习能力。熵权-BP神经网络结构如图3所示：



图3 熵权-BP神经网络结构

参照上图，熵权-BP神经网络建模过程：

Step1:第1层为输入层，其神经元的个数l为信息系统安全性评价指标体系的14个评价指标u；

Step2:第2层为指标筛选层，通过构造评价指标矩阵采用熵权法计算第1层评价指标的权重，并选取其中权重较大的评价指标x；

Step3:第3层为传递层，将第2层选取的n个权重较大的评价指标x传递到BP神经网络中；

Step4:第4层为隐含层，主要处理评价指标到评价结果之间的非线性的映射关系，隐含层神经元y的个数的选取是BP神经网络学习成功的关键，借助经验公式：

（12）

式中：n表示传递层神经元个数，β∈[1,10]的常数。

Step5:第5层是输出层，根据实际需要只设定一个神经元z，代表信息系统安全性评价结果。

此时，熵权-BP神经网络建模已完成，接下里就是第3层到第4层之间的权值wij与第4层到第5层之间的权值wj的学习过程。

## 2.4 BP神经网络学习算法：

如图3所示，传递层神经元用i表示，隐含层神经元用j表示。给定h组样本的输入和输出分别为：





BP神经网络学习过程：

1）随机选取BP神经网络的偏置值以及权值wij和wj；

2）BP神经网络隐含层和输出层样本p的正向传递：

 （6）

 （7）

其中f为非线性变换函数Sigmoid，θj表示隐含层神经元j(j-1,2,…,m)的偏置值，θ表示输出层神经元的偏置值；

3）使用α-η联合算法，反向修正BP网络权值和偏置值：

（8）

（9）

（10）

式中，f’为对f求导，η∈(0,1)为学习速率，加快网络的收敛速度；α∈(0,1)为动量因子，避免BP网络陷入局部最优。

4）h组样本拟合误差函数为：









式中ε(0≤ε<1)为误差精度，T为网络学习最大次数，当网络误差达到精度要求或者达到学习的最大次数，BP网络学习过程结束，否则转向步骤2)，继续学习直到满足条件为止。

# 3 实例研究

## 3.1 数据来源

为检验熵权-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,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 | X1 | X2 | X3 | X4 | X5 | X6 |  | X14 | y |
| 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 |

## 3.2 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R,即



2）通过3.2节的计算步骤，得出各评价指标的权重W=(w1, w2,…, w14)：

W=(0.17,0.008,0.006,0.24,0.02,0.03,0.17,0.1,0.04,0.01,0.05,0.09,0.08,0.06)

按评价指标重要度排序后W’：

W’=(0.24,0.17,0.17,0.1,0.09,0.08,0.06,0.05,0.04,0.03,0.02,0.01,0.006,0.008)

权重小于0.05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。

## 3.4 熵权-BP神经网络模型训练

设定样本训练的学习率η为0.5，动量因子α为0.3，最大学习次数T为6000，拟合误差ε为0.0001。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，此时熵权-BP模型已达到拟合误差要求。为了比较熵权-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。

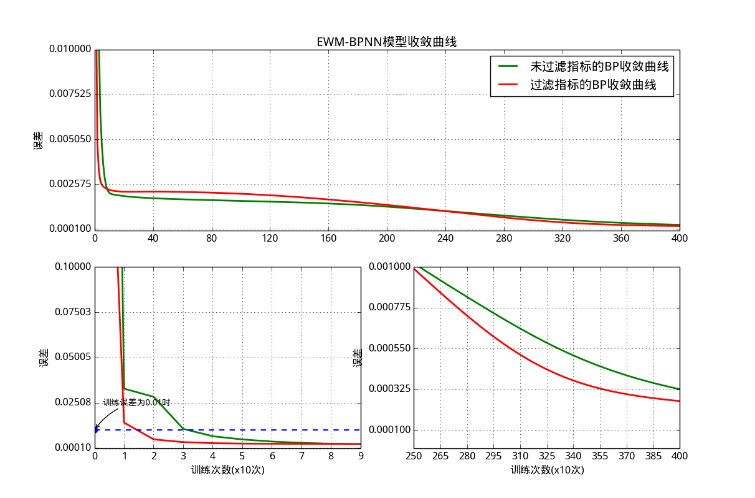


图4 BP神经网络训练集的收敛过程

由图4可知，熵权-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，熵权-BPNN的收敛速度也明显比BP神经网络快，最终训练完成后熵权-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，熵权-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。

## 3.5 熵权-BP神经网络模型检验

将测试样本分别输入到BPNN、AHP-BP和熵权-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）使用熵权-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）熵权-BP模型的评价结果精度高于AHP-BP模型的评价结果，这是因为熵权-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。

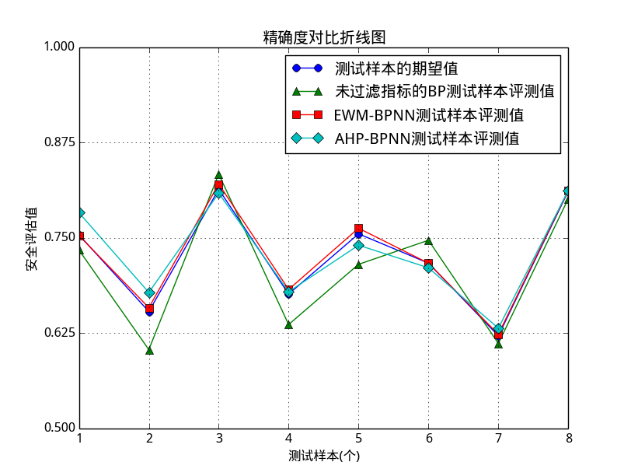


图5 各模型的网络输出值和期望值之间的对比

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 |
| 熵权-BP | 96.14 |

# 4 结束语

本文将熵权法和神经网络相结合，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，熵权-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. 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.
3. 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.
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. 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)
6. 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.
7. 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.
8. 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.
9. 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.
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. 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.
13. 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.
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.

联系方式：  
邮编：300300  
地址： 天津市东丽区津北公路2898号中国民航大学北院48号信箱  
联系人：王双  
联系电话：13612183687