**基于熵权法和神经网络的信息系统安全性评价**

**摘要：**

为了适应当前复杂信息系统的安全性评价，提出一种基于熵权法和神经网络相结合的信息系统安全性评价模型（EWM-BPNN），采用熵权法客观的计算样本评价指标数据的权重，并筛选出对综合评价结果有重要影响的指标作为BP神经网络的输入，采用BP神经网络构建信息系统安全评价模型。实验结果表明，EWM-BPNN模型不仅提高了指标质量，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，为信息系统安全性评价提供了一种新的方法。

**1 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，会对企业及国家安全带来灾难性的后果。比如，2014年8月1日，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。2015年12月23日,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的研究热点问题[1]，如何准确地评价信息系统安全状况具有重要的研究价值和现实意义。

信息系统安全性评价主要为定性分析和定量分析两种方式[2]，由于影响信息系统的安全因素难以量化，大多数学者采用定性与定量相结合的方法进行综合评价，如层次分析法、模糊理论、故障树等[3,5]。虽然这些传统的方法有较深的理论基础，但均存在操作复杂、计算量大、精度低、主观性强等问题。而且传统评价模型只能得到信息系统在某一时刻的安全状态，不能综合比较信息系统在不同时刻的安全状态[6]。针对上述问题，D-S证据理论、BP神经网络、RBF神经网络、、支持向量机、贝叶斯网络等人工智能算法已广泛应用在信息系统安全评价领域[7,9]。相比其他人工智能算法，BP神经网络具有自学习、自适应能力，可以有效的解决评价指标到评价结果的高度非线性映射关系，训练好的BP模型能评价具有类似特点的不同安全等级信息系统。Hoh P I等人通过软件风险管理技术，提出综合的定量风险分析模型，可以有效分析安全风险[10]； Ana Paula等人通过模拟事故发生场景采用事件树分析其潜在的威胁，但其鲁棒性较差[11]；张利等人，提出了4种改进的具有代表性的信息安全风险评估方法并对这4种方法的优缺点进行了对比分析[12]；马丽仪采用模糊神经网络模型评价信息系统的安全风险，在模糊处理时隶属度矩阵和权重的确定均采用专家评判，客观性较差，在一定程度上影响整个模型的实用性[13]；赵保华采用层次分析法和神经网络构建信息系统安全评价模型，在用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[14]。

为了解决以上问题，并进一步提高信息系统的安全评价精度，提出基于熵权法（[Entropy](javascript:void(0);) [Weight](javascript:void(0);) [Method](javascript:void(0);)，EWM）和BP神经网络（BP Neural Network，BPNN）相结合的信息系统安全性评价模型，实验结果表明,EWM-BPNN评价模型提高了信息系统安全的评价精确度，并为信息系统的安全评价实现自动化。

**2 EWM-BPNN信息系统安全性评价模型**

**2.1构建信息系统安全性评价指标体系**

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。申时凯等只是单一的从技术层面选取了6信息安全风险因素，评价指标覆盖范围不全面，导致评价结果可信度不高[15]。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[16]。信息系统安全性评价指标体系的递阶层次结构如图1所示。

**图1 信息系统安全性评价指标体系结构图**

**2.2 熵权法筛选重要评价指标**

在信息系统安全评价时，各评价指标在整个评价体系中的重要程度不同，赋予评价指标的权重极为重要，权重是否科学、合理直接影响评价结果的准确性。

评价指标权重的确定主要是主观和客观两种方式。主观权重一般多凭主观经验赋值，在一定程度上对评价结果有参考价值，但有时会严重扭曲客观实际，使评价结果严重失真。由于不同信息系统的评价专家可能不同，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。而客观确定权重是依据实际数据通过计算得到，不仅降低了人为因素，而且结果真实、可信。熵权法根据各评价指标的变异度，利用信息熵计算出各评价指标的权重，从而得到较为客观的评价指标权重。

本文利用已完成安全评价的信息系统样本数据，构建评价指标矩阵，计算得到各评价指标权重，通过权重排序，剔除不重要指标，筛选重要指标作为BP神经网络模型的输入，优化BP模型的输入，降低计算复杂度，提高拟合效率和信息系统安全评价准确率。

**2.2.1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评价指标的实际数值。

**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，计算各评价指标在不同信息系统的比重。

（2）

代表第个评价指标下第个信息系统的评价指标值的比重。

（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越大，其权重也应该越大。若某评价指标的熵越大，说明该指标值的变异程度越小，在评价中几乎不起作用。

（3）用对各评价指标的熵值进行归一化处理：

 （4）

（4）当熵值最大时，评价指标在综合评价中作用越小，熵值和评价指标成反

比，故引入差异系数 来正向度量评价指标的熵权，得到评价指标的权重：

（5）

式中，。

**2.2.4 筛选重要性指标**

通过以上步骤得到所有评价指标权重并对其按重要度进行排序，剔除在信息系统安全评价中几乎不起作用的评价指标，筛选出重要的评价指标作为BP神经网络的输入。

**2.3 BP神经网络模型**

**2.3.1 BP神经网络结构**

BP神经网络简称为BP网络。BP网络包括输入层，输出层，隐含层，共三部分，BP网络每一层都有若干个神经元。BP网络是一个“多输入-多输出”的高度非线性的映射关系。目前被广泛用在综合评价领域，典型的三层BP网络拓扑结构如图2所示：

**图2 BP神经网络结果拓扑图**

**2.3.2 BP算法**

BP算法，即误差反向传递学习算法。BP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，BP网络模型已训练好。如2.3.1节的BP网络结构图，输入层、隐含层和输出层神经元分别用，，表示。

BP神经网络算法的基本步骤：

（1）初始化参数，随机给定BP神经网络的权值和阈值；

（2）给定组样本的输入和输出分别为、；

（3）BP神经网络隐含层和输出层正向传递；

隐含层和输出层神经元的输入和输出之间的非线性关系用Sigmoid函数，即（7）

隐含层样本的输入和输出分别为：

 （8）

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：

 （9）

其中表示输出层神经元的偏置值。

（4）使用联合算法，反向修正BP网络权值和偏置值：

（10）

（11）

（12）

式中，为学习速率，为了加快网络的收敛速度。为动量因子，避免陷入局部最优。

**（5）**组样本拟合误差函数：

（13）

式中为期望输出值，为网络实际输出值，是误差精度，是很大的整数，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。

**3. EWM-BP神经网络信息系统评价步骤**

**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** 通过评价指标权重的排序，筛选出对综合评价结果有重要影响的指标；

**步骤4**将筛选出的重要评价指标作为BP网络的输入，即为BP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EWM-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。

EWM-BP神经网络信息系统评价流程如3图所示：

**图3 EWM-BP神经网络信息系统评价模型**

**4 实验仿真**

**4.1 数据样本来源**

为检验EWM-BP神经网络信息系统安全性评价模型的性能，对其进行仿真实验。收集民航已完成安全评价的30个信息系统作为样本数据,如表1。将30个样本分为20个训练集和10个测试集两部分。

**表1 信息系统安全评价数据**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 样本 |  |  |  |  |  |  |  |  |  |
| 1 | 68.4 | 59.43 | 68.43 | 62.7 | 85.08 | 74.68 |  | 80.97 | 0.754 |
| 2 | 75.4 | 65.2 | 75.67 | 82.84 | 84.48 | 73.32 |  | 84.62 | 0.8136 |
| 3 | 61.64 | 62.59 | 75.11 | 65.8 | 79.68 | 94.68 |  | 87.37 | 0.6726 |
| 4 | 69.28 | 59.66 | 72.35 | 60.27 | 86.08 | 80.68 |  | 76.87 | 0.7552 |
| 5 | 70.16 | 73.53 | 75.36 | 72.66 | 77.32 | 81.6 |  | 68.25 | 0.7168 |
| 6 | 91.93 | 67.29 | 63.4 | 82.24 | 91.7 | 100 |  | 78/72 | 0.8315 |
|  |  |  |  |  |  |  |  |  |  |
| 30 | 78.74 | 75 | 61.6 | 62.17 | 81.16 | 94.68 |  | 84.85 | 0.7657 |

**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2.2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0.05以下的评价指标具体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个重要的评价指标作为BP神经网络模型的输入。

**4.3数据预处理**

该企业评价指标使用是百分制，为降低BP神经网络计算复杂度，提高训练效率，对其归一化处理，如下：

（14）

其中为评价指标的最小取值，为评价指标的最大取值。

**4.4 EWM-BP神经网络建模**

EWM-BP神经网络共有三层，输入层神经元的数量为通过熵权法筛选出重要的评价指标个数，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点，借助经验公式：

（6）

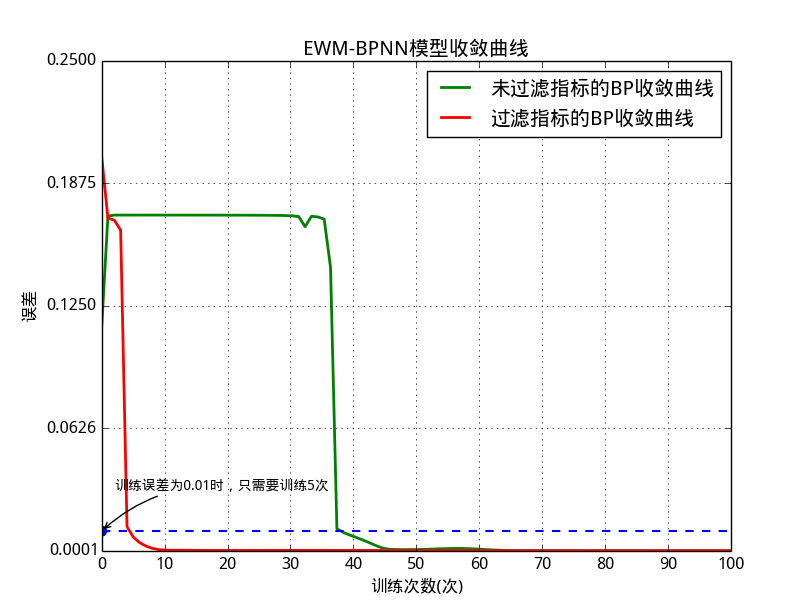
式中：表示输入层神经元个数，的常数。

设定样本训练的学习率为XX,动量因子为XX,拟合误差为。

**4.5 EWM-BP神经网络信息系统安全评价模型训练和检验**

**4.5.1 EWM-BP神经网络信息系统安全评价模型训练**

将归一化后的重要评价指标训练集输入到BP模型中学习训练，在训练125次时拟合误差就达到了0.0001，此时的BP神经网络模型已达到实验要求，训练EWM-BP神经网络训练集的收敛曲线如图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神经网络模型~~只需要迭代70次就可以达到0.002的误差，~~当训练次数达到125次时，误差减小为0.0001，而没有使用熵权法筛选评价指标的BP模型在迭代次数为150次时，其误差为0.17117，当训练次数达到XXX次时，误差减小为0.0001。

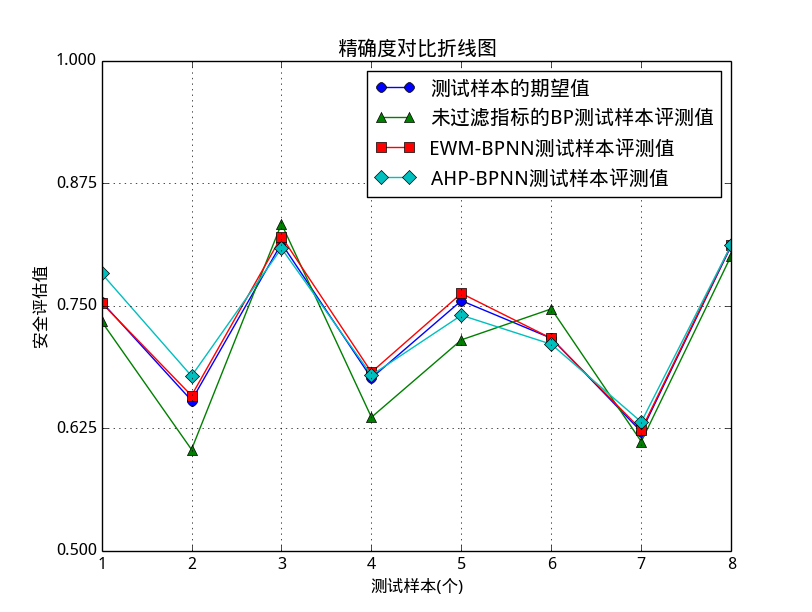
实验结果表明，EWM-BP神经网络模型通过对筛选出8个重要的评价指标样本数据进行训练，可以有效的减少迭代次数，并快速收敛到设定的误差，既可以提高BP神经网络的训练速度，又能保留重要指标的特征，减少非重要指标对BP神经网络模型的干扰，保证BP神经网络评价模型的准确度。

**4.5.2 EWM-BP神经网络信息系统安全评价模型检验**

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**

**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-BPNN | 93.25 |
| EWM-BPNN | 96.14 |

**5 结束语**

针对当前复杂的信息系统安全评价问题，提出一种基于EWM-BPNN的信息系统安全评价模型，并通过实验仿真验证模型。仿真结果表明，EWM-BPNN通过样本采用熵权法计算评价指标的权重，筛选出重要指标，提高评价指标质量，同时结合具有高度非线性处理能力的BP神经网络，不仅提高了信息系统安全评价的精度，而且简化了评价过程，非常适合应用到信息系统安全评价领域。

**参考文献：**

[1] 陈颂，王光伟，刘欣宇，等.信息系统安全风险评估研究[J]．通信技术，

2012，45(1)：128-130.

[2] 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）

[3] 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．

[4] 章文辉，杜百川，杨盈昀．模糊层次分析法在广播电视信息安全保障评价指标体系中的应用研究[J]．电子学报，2008，36(10)：2060-2064.

[5] Rosqvist, T., Molarius, R., Virta, H., & Perrels, A. (2013). Event tree analysis for floodprotection-an exploratory study in Finland. Reliability Engineering & SystemSafety, 112, 1–7.

[6] 张俊林.网络安全风险动态评估方法研究[J] 计算机仿真,2016:33(10)

[7] 党德鹏，孟真．基于支持向量机的信息安全风险评估[J]．华中科技大学学报：自然科学版，2012，38(3):46-49．

[8] 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):

[9] 付钰, 吴晓平, 严承华. 基于贝叶斯网络的信息安全风险评估方法[J].

武汉大学学报: 理学版, 2006, 52(5): 631-634.

[10]Hoh P I,Kim Y G,Lee T,et al. A security risk analysis model for information systems[C]//Baik D K,Lecture Notes in Computer Science. New York,USA：Springer Verlag Inc,2005:505-513.

[11] Ana Paula Henriques de Gusmão∗, Lúcio Camara e Silva, Maisa Mendonc. Information security risk analysis model using fuzzy decision theory[J]. International Journal of Information Management,2016,36(1): 25–34.

[12] 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．

[13] 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) : 267—273.

[14] 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)：

[15]申时凯，佘玉梅.模糊神经网络在信息安全风险评估中的应用[J] 计算机仿真，2011：28(10)

[16] 中华人民共和国国家质量监督检验检疫总局，中国国家标准化管理委员会．信息安全技术 信息系统安全等级保护测评要求：GB/T 28448-2012

[17] 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）, 2014，54(1)： 35-43，53.