**基于熵权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估，提出一种熵权法和神经网络的信息系统安全性评估模型，结果表明，此模型可以客观的剔除掉不重要的评价指标，并有效解决了信息系统的不确定性和动态变化性，提高了信息系统安全性的评估精度。

0 引言

随着各行业对信息系统的依赖度日益增加，信息系统安全事故也频繁发生。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，信息系统的安全问题成为当前的研究热点。

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，其中定性评价不直观反映信息系统安全等级，目前大多学者采用定量的方法评价信息系统的安全性，其中包括层次分析法、模糊理论、故障树等。这些方法均存在操作复杂、计算量大、精度低、主观性较强等问题，已不适合用来评价当前复杂的信息系统。目前，D-S证据理论、BP神经网络、遗传学等人工智能算法已广泛引用于评价领域。其中BP神经网络具有高度处理非线性数据关系的能力、自学习、高拟合等，可以解决信息系统的不确定性。马丽仪采用模糊BP神经网络模型评价信息系统的安全风险并得到很好的效果，但其样本数据均采用专家评判，客观性较差，影响评价结果。赵保华采用层次分析法和神经网络得到稳定的评估模型，但用层次分析法剔除无用指标，主观性、随意性较大。熵权法利用实际样本数据，通过指标变异度客观的确定指标权重，避免了人为因素。所以，本文提出了基于熵权法和BP神经网络的信息系统安全性评价模型，实验结果表明熵权法和BP神经网络评价模型可以充分利用样本数据的有效信息，提高了信息系统安全性评估精确度。

1**信息系统安全评价指标体系**

评估信息系统安全性首先是构建评价指标，即信息系统安全影响因素。信息系统的复杂、多样化，使得提取指标十分困难。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。申时凯等只是单一从技术层面选取指标，指标覆盖范围不全面。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》，从不同的方面对信息系统进行全方位的评价，主要分为技术方面和管理方面两大类。信息系统全性指标体系的递阶层次结构如图所示。



信息系统安全性评价指标体系结构图

2 熵权法筛选评价指标

在信息系统安全评价时，各指标在整个评价体系中的重要程度不通，赋予指标正确的权重极为重要，权重是否科学、合理直接影响评价结果的准确性。评价指标的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。赵保华用AHP确定指标权重剔除不重要指标，主观性较强，可能会将重要指标误剔，影响评价结果准确性。由于不同信息系统评价专家可能不同，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。客观确定权重是依据实际数据分析得到，本文利用已收集的信息系统指标实际评价样本，通过信息熵客观的分析同一指标在不同信息系统的变异程度，确定各指标重要度，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，从而降低BP模型的输入维数，降低计算复杂度，提高拟合效率和评价准确率。

2.1 熵权确定指标权重

（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，与个评价指标形成矩阵：



其中为第个信息系统第个指标的实际值。

指标有熵值：



由熵的极值性可知，若某指标的熵越小，说明该指标值的变异程度越大 ，提供的信息量越多，在综合评价中起的作用越大，其权重应该越大。若某指标的熵越大，说明该指标值的变异程度越小，在评价中几乎不起作用，应将其剔除。

（2）用对信息熵进行归一化处理

当熵值最大时，指标在综合评价中作用越小，熵值和指标成反比，用来正向度量指标的熵权。

（3）计算指标的权重：



式中，。

2.2 重要性指标筛选

利用2.1步骤计算全部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指标，保留重要指标。

3 BP神经网络的信息系统风险评价模型

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网络结构的特点是上下层神经元之间全连接，同一层神经，元之间无连接，各层神经元之间无反馈连接,它的输入和输出是一个“多输入-多输出”，且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网络拓扑结构如图所示：



3.2 BP算法

BP算法，即反向传递学习算法。BP网络学习包括输入信号的正向传播和误差信号反向传播两个阶段。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3.1节的BP网络结构图，输入层和隐含层神经元个数分别为，。其中为评价指标的个数。隐含层个数有实验经验得出。结合实际需要，输出层只设一个神经元，代表信息系统安全值。

令组的输入、输出样本式如下所示：



1. 输入层神经元的输入和输出相同，即为评价指标值。
2. BP网络中，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



样本隐含层神经元的输入和输出分别为：





其中表示隐含层神经元的偏置值。

样本输出层神经元的输入和输出分别为：

其中表示输出层神经元的偏置值。

1. 结合联合算法，神经元权值和偏置值按下列规则调节















式中，为自适应学习速率，是为了加快网络的收敛速度。为动量因子，避免陷入局部最优。

1. 组样本拟合误差函数：

式中为网络的实际输出值，为输出层的网络输出值。

是误差精度，，或者达到循环要求的最大次数，即，是很大的整数，网络训练的最大次数。

3.3 熵权-BP神经网络信息系统评价模型

**步骤1** 参考XXX文献，结合访谈评价信息系统的专家，以及结合《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** 通过收集已评价的信息系统样本数据，利用熵权法对数据进行分析，得到评价指标的权重

**步骤 3** 通过评价指标的排序，筛选出对综合评价结果有重要影响的指标。

**步骤 4** 对样本数据进行归一化处理，方便BP网络模型的训练。

**步骤 5** 熵权法筛选出的重要指标作为BP网络的输入，即为BP网络输入层神经元的个数。BP网络的输出层设定一个神经元，即为信息系统安全性输出值。隐含层个数根据公式确定。



表示输入层神经元个数，的常数。

**步骤 6** 用采集的样本训练BP网络，得到稳定的信息系统安全评价模型。

**步骤 7** 用测试样本对训练好的信息系统安全评价模型进行检测**，**并分析其精确度。

　本文采用熵权-BP神经网络结合的方式对信息系统安全性评价，通过采集实际指标数据，运用信息熵理论分析各个指标在信息系统综合评价中价值，剔除不重要指标。将剩余指标经归一化处理作为BP模型的输入层神经元。根据《信息安全技术信息系统安全等级保护定级指南》将信息系统安全等级划分，分为5个评价等级，评价等级论域。各等级区间取值分别为：为低，为较低，，为中等，为较高，为高。熵权法-BP模糊神经网络信息系统安全评价过程。



熵权-BP神经网络信息系统评价模型

4 实验仿真

4.1 数据样本来源及预处理

为检验熵权-BP神经网络信息系统安全评价模型的性能，对其进行仿真实验。收集民航已完成测评的100个信息系统作为样本数据,如表X。将100个样本分为80个训练集和20个测试集两部分

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 | 89.02 | 80.97 | 0.754 |
| 2 | 75.4 | 65.2 | 75.67 | 82.84 | 84.48 | 73.32 | 88.46 | 84.62 | 0.8136 |
| 3 | 61.64 | 62.59 | 75.11 | 65.8 | 79.68 | 94.68 | 80.93 | 87.37 | 0.6726 |
| 4 | 69.28 | 59.66 | 72.35 | 60.27 | 86.08 | 80.68 | 91.12 | 76.87 | 0.7552 |
| 5 | 70.16 | 73.53 | 75.36 | 72.66 | 77.32 | 81.6 | 66.4 | 68.25 | 0.7168 |
| 6 |  |  |  |  |  |  |  |  |  |

该企业评价指标是百分制，为了适应BP神经网络激励函数Sigmoid对之间的数字比较敏感，故其归一化处理。

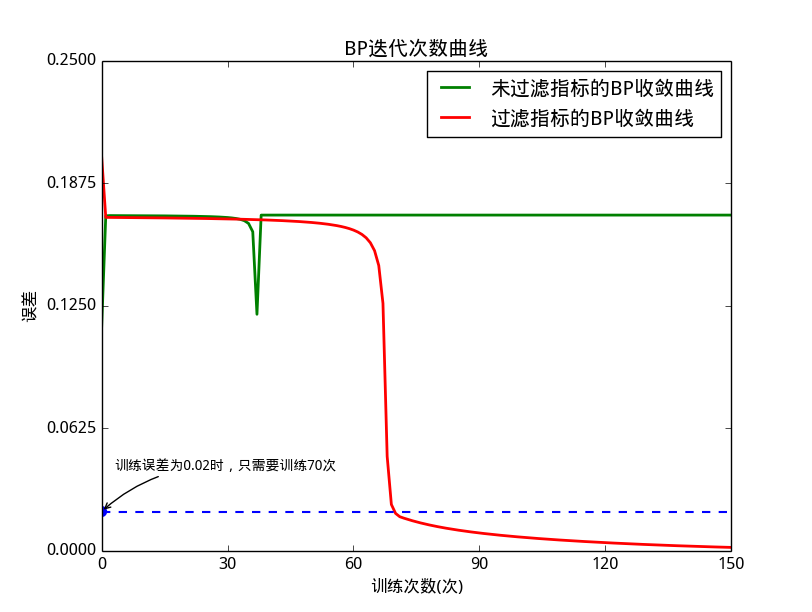


其中为指标的最小取值，为指标的最大取值。

4.2 熵权-BP神经网络信息系统安全评价模型训练和检测

（1）熵权-BP神经网络信息系统安全评价模型训练

将经过熵权法筛选的指标作为BP网络的输入层。将训练集输入到BP模型训练，在训练125次时精度达到0.0001。此时的BP神经网络模型已达到实验要求，训练BP模型的收敛曲线如图XX所示。



其中，X线是熵权-BP神经网络模型训练次数与误差关系曲线，Y线是未过滤指标的BP神经网络模型训练次数与误差关系曲线，由图X可知，只需要迭代70次熵权-BP神经网络模型就可以达到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。

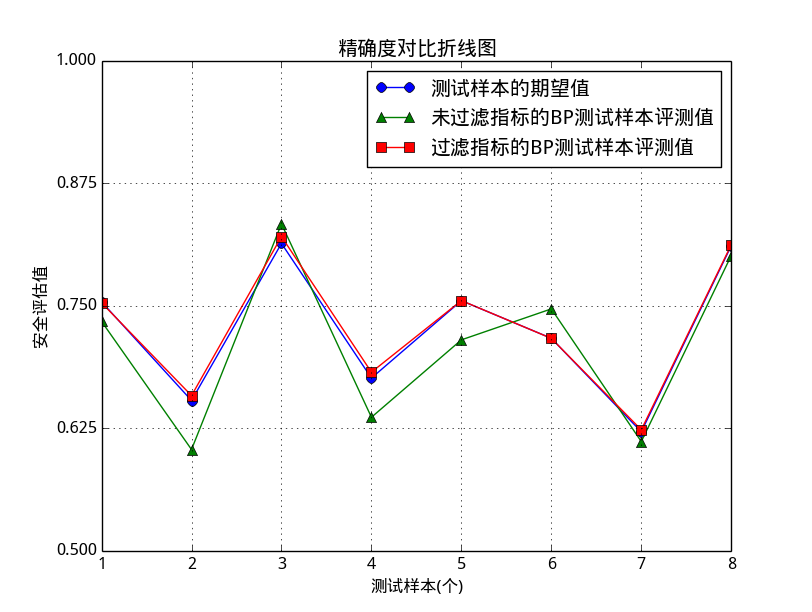
实验结果表明，使用熵权-BP神经网络模型通过筛选出重要的指标进行模型训练，可以有效的减少模型迭代次数，并快速收敛到实验要求误差，既可以提高模型训练速度，又保留重要指标的特征，减少非重要指标对模型的干扰，保证模型的准确度。

（2）熵权-BP神经网络信息系统安全评价模型检测

将测试样本分别输入到BPNN、EWM-BP、AHP-BP中得到信息系统的安全评价期望值与评估值的对比关系图如图X所示，同时，表XX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估结果比直接使用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，过滤掉影响评估结果的次要指标，使得模型在反向传播过程中更加逼近真实权重，从而使得最终评价结果更接近期望值。

2.使用EWM-BP模型得到的评估结果优于使用AHP-BP模型得到的评估结果，这是因为EWM-BP模型通过样本数据计算指标的信息熵，更加客观的反映指标权重，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，使得指标的筛选不稳定，模型效果受主观因素影响较大。



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

5 结束语