神经网络在确定关联规则挖掘算法权值中的应用研究*

李彤岩，李兴明
(电子科技大学 宽带光纤传输与通信网技术教育部重点实验室，成都 610054)

摘　要：提出了运用神经网络确定权值的方法，将网络告警信息的三个主要属性作为神经网络的输入，通过样本的训练来确定神经网络的连接权，从而识别网络告警的权值。这种权值确定法既体现了专家的经验知识，又能够随着网络拓扑的变化更新连接权。建模及仿真结果表明，与其他权值确定方法相比，神经网络方法更加实用和有效。

关键词：加权关联规则；通信网告警相关性分析；神经网络；连接权
中图分类号：TP301　　**文献标志码**：A　　**文章编号**：1001-3695(2008)05-1440-02

Study of neural network applied to weighted association rules mining

LI Tong-yan, LI Xing-ming
(Key Laboratory of Broadband Optical Fiber Transmission & Communication Networks of Ministry of Education, University of Electronic Science & Technology of China, Chengdu 610054, China)

Abstract: A novel method which used neural network to identify the alarm weight was proposed. The neural network had three inputs with the key elements which reflected the importance of the telecommunication alarm. After the course of sample training, the link weight would be confirmed. The weight of the neural network might reflect the knowledge of the experts and also could be changed automatically with the different items from the inputs. Modeling and simulation study indicates that compared with other methods of measuring alarm weight, the neural network method has more advantages.

Key words: weighted association rules; telecommunication alarm correlation analysis; neural network; link weight

随着通信网的迅速发展和通信业务的扩展，网络的故障管理也就显得至关重要。利用数据挖掘技术中的关联规则挖掘对网络告警信息进行相关性分析，从而查找故障的根源是目前研究的热点。关联规则挖掘算法[1~3]最早是由 Agrawal 等人提出来，其算法思想是要找出不同数据之间的关联性。这正适合通信网告警相关性分析的研究。现有的关联规则挖掘算法大多是将数据库中的记录看成平等的，即各个项目具有相同的权值，然而通信网告警数据的特点是告警通常由很多属性构成，告警级别是其中表征告警严重程度的一个主要度量；告警设备所处的节点度数不同，对网络产生的影响也会不同。在关联规则挖掘时不能简单地将各项告警记录之间的关系看做是平等的。为了解决这个问题，可以综合考虑影响告警属性的各种因素，为告警信息分配相应的权值，采用加权关联规则挖掘算法来进行通信网告警相关性分析。加入了权值，就可以挖掘出更多重要信息，而去除一些并不重要的告警信息，从而增加挖掘的效率，减少挖掘的数据空间。权值的分析本身就是一个难点，而且通信网是一个拓扑结构动态变化的网络，告警的权值要能够反映网络动态变化的特性。

一般告警权值是依靠网络专家的经验来确定的，但是在现代通信网的条件下，告警数量成千上万，网络变化错综复杂，单单依靠人力是不可能很好地解决，所以就迫切需要一种既能够体现专家的经验知识又能够适应复杂的通信网特点的权值识别方法。文献[4]提出了一种加权关联规则挖掘算法的思想。其算法是基于购物篮分析的，项目集中各项权值可以由物品的价格和利润等很直观的因素决定。但是这种方法并不适合通信网告警权值的确定。因为其权值是由很多因素决定的，有些因素并不好定量分析，而且各个因素对网络的影响程度也不相同，不能直接使用已有的加权关联规则挖掘算法。文献[5,6]提出了层次分析法来确定告警权值，但是其中的参考因素太客观，不能很好地处理告警权值，而且在网络拓扑发生变化时必须重新确定权值，不能够参照以前的经验知识，就使得权值的更新变得非常复杂。

本文采用神经网络的方法将影响网络告警的因素作为神经网络的输入。通过样本学习来训练连接权，使得神经网络的连接权可以存储网络的信息。其中既包括告警属性这样的客观因素，又包括专家经验这样的主观因素，能够快速有效地确定权值。如果网络的拓扑信息发生变化，还可以对连接权进行更新，通过更新算法的研究，使得告警可以充分体现网络动态变化的特点，具有很强的实时性。

1　应用神经网络确定告警权值

神经网络的基本思想是从仿生学的角度模拟人脑神经系统的运作方式，使机器具有人脑那样的感知、学习和推理能力。在神经网络中，记忆的信息存储在连接权上，通过反复训练，可以在下一次输入信息时，通过连接刺激相应的神经元，从而达到自动识别的目的。虽然在数据挖掘中对权值的分析已有了一些研究成果，但是自动识别的效果都不是很理想。本文引入

收稿日期：2007-04-19；修回日期：2007-07-26　　基金项目：国家自然科学基金资助项目(60572091)
作者简介：李彤岩(1980-)，女，河南洛阳人，博士研究生，主要研究方向为通信告警相关性分析、网络故障诊断与定位、分布式网络管理等(sunny60138800@yahoo.com.cn)；李兴明(1956-)，男，云南人，教授，博导，主要研究方向为现代通信网、光网络与自动光交换、电信网管理等.

神经网络中的一神经元前馈模型即感知机模型来对告警权值进行学习和识别,最后通过仿真来说明运用神经网络确定告警权值的优势。

本文设计中影响网络告警权值的因素有三个,神经网络的输出是告警的权值,所以只需要一个神经元就可以完成识别功能,可以用神经网络中的感知机模型来完成(图 1)。

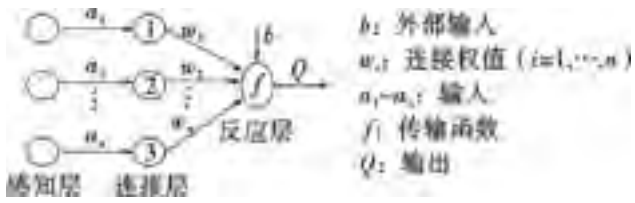

图 1　单层感知机的拓扑结构

感知机模型是一种具有权值元素的前馈网络,通常由感知层、连接层和反应层构成。其中感知层由输入元素组成。感知层至连接层的连接权值初始值固定为 1,在学习过程中这个权值不变化;连接层至反应层的连接权是通过学习来确定的。通常这些连接权在初始化时取[- 1, 1] 区间内的随机值,在学习阶段根据学习规则并利用输入/输出模式不断调整,使之在一定的输入时有期望的输出。

在这个模型结构中需要确定三个参数,即输入端的定量描述、神经网络的连接权和神经网络的传输函数。

1) 神经网络的向量表示

神经网络输入: $P_1 = \begin{bmatrix} a_{11} \\ a_{12} \\ \cdots \\ a_{1m} \end{bmatrix}, \cdots, P_n = \begin{bmatrix} a_{n1} \\ a_{n2} \\ \cdots \\ a_{nm} \end{bmatrix}$。

其中 $a_{n\times m} = \begin{bmatrix} a_{11} \cdots a_{1m} \\ \cdots \quad \cdots \\ a_{n1} \cdots a_{nm} \end{bmatrix} = \begin{bmatrix} P_1^{\mathrm{T}} \\ \cdots \\ P_n^{\mathrm{T}} \end{bmatrix}$。

神经网络连接权: $W_{m\times 1} = \begin{bmatrix} w_1 \\ \cdots \\ w_m \end{bmatrix}$

神经网络的净输入: $n = a \times W + b = \begin{bmatrix} P_1^{\mathrm{T}} \\ \cdots \\ P_n^{\mathrm{T}} \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \cdots \\ w_m \end{bmatrix} + b$

神经网络输出: $Q = f(n) = f(a \times W + b) = f(a_{n\times m} \times W_{m\times 1} + b) = f\left(\begin{bmatrix} P_1^{\mathrm{T}} \\ \cdots \\ P_n^{\mathrm{T}} \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \cdots \\ w_m \end{bmatrix} + b\right)$。其中: f 表示神经网络的传输函数。

2) 神经网络连接权的训练过程

连接权的训练过程分为以下几个步骤:

a) 选取样本数据$\{p_1, q_1\}, \cdots, \{p_n, q_n\}$。其中: $P_1 = (a_{11}, \cdots, a_{1m})^{\mathrm{T}}, \cdots, P_n = (a_{n1}, \cdots, a_{nm})^{\mathrm{T}}$($p$ 表示输入元素; q 表示输出元素)。

b) 设定连接权的初值$\{w_1^0, \cdots, w_n^0\}$。

c) 定义感知机的误差为 $e = q - Q$。其中: q 是样本值; Q 为感知机输出值。

d) 权值调整的规则为: $e > 0$, $w^{\mathrm{new}} = w^{\mathrm{old}} + p$; $e < 0$, $w^{\mathrm{new}} = w^{\mathrm{old}} - p$; $e = 0$, $w^{\mathrm{new}} = w^{\mathrm{old}}$。其中: p 表示样本的输入值; w^{old}表示调整前的权值; w^{new}表示调整后的权值。

通过对误差的判别来调整连接权值,若误差达到评价标准,调整过程结束,输出的权值就是已经训练好的权值;否则就返回重新评价。连接权的训练流程如图 2 所示。

2　仿真及分析

仿真输入预处理后的通信网告警数据,其告警信息属性由多个因素组成,选用单神经元的前馈神经网络模型来确定告警权值(图 3)。输入分别选用对通信网影响最大的包括告警设备的节点度数、告警级别、告警类型在内的三个属性;输出为每条告警信息的权值。通过对专家经验知识即样本的训练,得到神经网络的连接权值,建立好神经网络模型后,就可以判定告警权值了。

图 2　神经网络连接权值的训练流程

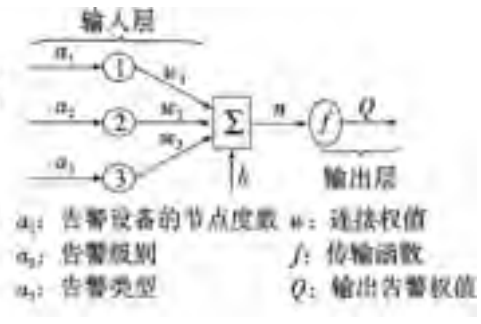

图 3　神经网络设计图

告警相关性分析确定权值的神经网络设计过程中,确定传输函数是最后也是关键的一步。在本文设计中,根据告警数据的特点选取传输函数为饱和线性函数:

$$f = \begin{cases} 0 & n < 0 \\ n & 0 \le n \le 1 \\ 1 & n > 1 \end{cases}$$

仿真选取两个典型网络 NSFNET 和 CERNET 的 10 ~50 条告警的典型权值对神经网络的连接权进行训练。最后的告警连接权值趋于稳定,表明训练完成(图 4)。

分别利用神经网络方法和层次分析法对事务数据库中的告警项进行加权,然后采用加权 Apriori 算法[4] 挖掘加权频繁项集。程序是在 AMD Processor 2800 +, 内存 512 MB, Windows XP 环境下用 Java 语言实现的,使用 Java 5.0 对程序进行编译。图 5 是两种加权方式下不同的最小加权支持度产生的加权频繁项目集个数的比较。

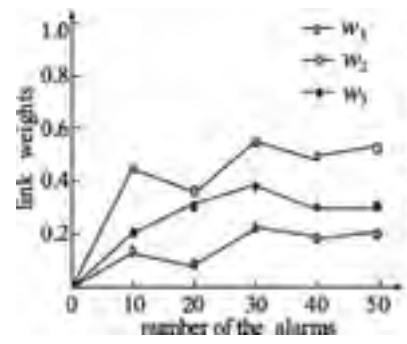

图 4　神经网络连接权值调整

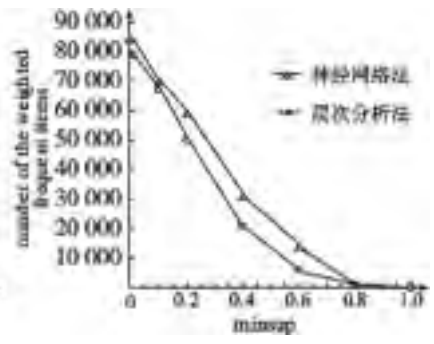

图 5　两种权值方法在关联规则挖掘中的应用比较

从仿真结果可以看出:随着最小加权支持度的增加,神经网络方式下加权频繁项目集个数下降得更缓慢。因为这种加权方式可以较好地参照专家的经验知识,又能够体现不同网络节点度的告警信息的差别,所以在不同的最小加权支持度下能挖掘出加权频繁项集,比层次分析法的识别性要强很多。

3　结束语

加权关联规则挖掘是适合于通信网告警相关性研究的一个重要的方法,告警权值的确定关系到挖掘的效果。本文通过理论分析,采用了前馈神经网络模型来确定通信网的告警权值,并且根据告警信息的特点对神经网络进行 (下转第 1449 页)

动智能决策分析模块中的实例检索引擎, 并根据实例库中复杂情况或事故类型的实例, 基于 CBR 技术进行决策分析, 得出其处理方案、预防措施以及专家分析处理意见等。从 “结果输出 ”按钮可查看其详细分析报告(报告为 Word 文件格式) 。

图 3　复杂情况与事故类型诊断参数录入界面

图 4　决策分析人机交互界面

运行系统的复杂情况与事故预防、处理指导手册, 可获得各种复杂情况与事故预防、处理知识, 包括复杂情况或事故发生的原因、发生时的特征、预防措施、处理方法以及处理流程等(图 5) ; 运行实例查询功能, 可从实例库中查询各油田已发生的复杂或事故实例, 查询结果示例如图 6 所示。

图 5　复杂与事故预防、处理指导示例

5　结束语

基于实例推理是对人类形象思维的模拟, 适于专家知识难以形式化表达且已积累丰富实例的决策问题领域。本文基于实例推理技术开发的井下复杂情况与事故智能诊断和处理系统克服了传统专家系统中存在的知识获取瓶颈问题, 且具有自学习、决策效率高等优点。该系统包含了近千个复杂与事故实例, 可对井下 27 类复杂情况与事故进行诊断, 并提供其预防措施、处理方案以及专家分析意见等; 还提供了完整的有关复杂与事故预防、处理的知识手册。本系统也是笔者与中国石油长城钻井公司合作项目的一个子系统, 已在海外两个国家的多个钻井队投入了实际应用。

图 6　实例查询示例

参考文献:

[1]　王东, 刘怀亮, 徐国华. 案例推理在故障诊断系统中的应用[J] . 计算机工程, 2003, 29(12) : 10-12.

[2]　张荣梅, 徐序彦. 基于 CBR 的交通事故处理智能决策支持系统[J] . 计算机工程与应用, 2002, 38(2) : 247- 249.

[3]　蒋希文. 钻井事故与复杂问题[M] . 北京: 石油工业出版社, 2001.

[4]　杨勇, 胡书林, 陈开联. 钻井工程事故分析与预报[J] . 录井技术, 2003, 14(2) : 54-59.

[5]　BURKHAND L D. Similarity and distance in case-based reasoning [J] . Fundamental Informatics, 2001, 54(17) : 201-205.

[6]　SLONM T Y, SCHNEIDER M. Design issues in fuzzy case-based reasoning[J] . Fuzzy Sets and Systems, 2001, 117(59) : 251-267.

(上接第 1441 页) 建模。通过对实际网络告警数据典型样本值的训练, 得到了有稳定连接权值的神经网络模型。仿真结果表明, 应用神经网络来确定告警权值, 既能够体现专家的经验知识, 又可以使权值的确定充分体现告警的属性, 而且在网络拓扑变化时可以快速动态地确定和调整权值, 使得加权关联规则挖掘更加科学、有效。

参考文献:

[1]　GRAWAL R, IMIELINSKI T, SWAMI A. Database mining: a performance perspective[J] . IEEE Trans on Knowledge and Data Engineering, 1993, 5(6) : 914-925.

[2]　AGRAWAL R, SRIKANT R. Fast algorithms for mining association rules[C] / /Proc of the 20th VLDB Conf. 1994.

[3]　AGRAWAL R, MANNILA H, SRIKANT R, *et al*. Fast discovery of association rules[C] / /Proc of Knowledge Discovery and Data Mining. [S. l.] : MIT Press, 1996.

[4]　CAI C H, FU A W C, CHENG C H, *et al*. Mining association rules with weighted itemsets[D] . [S. l.] : Department of Computer Science and Engineering, 1998.

[5]　欧阳继红, 王仲佳, 刘大有. 具有动态加权特性的关联规则算法[J] . 吉林大学学报: 理学版, 2005, 43(3) : 314-319.

[6]　肖海林, 李兴明. 层次分析法在通信告警加权关联规则挖掘中的应用研究[J] . 电信科学, 2006, 22(11) : 36-39.