

文章编号:1001-9081(2005)10-2331-03

## 基于 Kohonen 网络的软件可靠性模型选择

吴 勤,侯朝桢,原菊梅

(北京理工大学 信息科学技术学院,北京 100081)

(wuqin-430@sohu.com)

**摘 要:**软件可靠性模型是软件可靠性工程的一个重要方面。现在还没有一个通用的模型,模型选择问题已成为模型研究的重点。运用聚类思想对软件可靠性模型的选择进行研究。在对软件失效数据进行编码的基础上,采用 Kohonen 神经网络对其进行聚类分析,从而实现了可靠性模型的选择。最后通过仿真测试,证明了此方法的有效性和可行性。

**关键词:**软件可靠性模型;Kohonen 网络;模型选择

**中图分类号:**TP311 **文献标识码:**A

## Selection of software reliability model based on Kohonen network

WU Qin, HOU Chao-zhen, YUAN Ju-mei

(School of Information Science and Technology, Beijing Institute of Technology, Beijing 100081, China)

**Abstract:** Software reliability model selection is a important part of software reliability engineering. Now there is no model which is suitable for all users. Moreover, the accuracy of reliability measure for using different model is different. Model selection has been an emphasis in the research of model. In this paper, the clustering analysis was used as a tool for software reliability selection. On the basis of Failure data coded, Kohonen nets were used as a tool for clustering analysis. Twenty data sets were used to demonstrate the approach and efficiency.

**Key words:** software reliability model; Kohonen net; model selection

### 0 引言

软件可靠性是软件工程和可靠性工程相结合所产生的一个新的分支,而软件可靠性模型是分析和评估软件可靠性的基础。软件可靠性模型应用的不一致性的根源在于模型建立的假设各不相同,而现有的可供实用的模型,其基础都建立在各个不同的统计假设之上,这些假设界定了具体模型关于故障行为的概率方式,同时也界定了每个模型的不同适应范围。对于软件可靠性模型选择的随意性束缚和限制了软件可靠性模型的应用。因此,目前还没有一种被证明是简单又实用的适合于用户的通用方法,现有的模型所表现出来的可靠性度量准则差异很大,如何进行模型的合理选择仍热是研究的热点<sup>[1]</sup>。

软件可靠性模型的选择是一个复杂的决策问题,度量者的个人经验在模型选择上占到很大成分。文献[2]中给选择可靠性模型提出一种思路:可以近似认为相似的失效数据选用同一种模型是合适的。把不同的软件失效数据送入分类器进行聚类,分为同一类的软件选用同一种模型。在 Elewa 的启发下,并参考了文献中的思想和编码方法,本文提出了一种用 Kohonen 网进行可靠性模型选择的方法。

### 1 用于聚类的 Kohonen 神经网络

#### 1.1 聚类选择模型的原理

对于给定的软件失效数据由于不同软件其失效数据的个数不同,无法直接用失效数据进行聚类,所以首先要对此数据

进行编码。通常是用待选择的可靠性模型对此失效数据进行编码。在编码向量空间中,每一失效数据经编码后代表了此高维空间中的一点,表示  $N$  个失效数据集的  $N$  个点组成了指定的数据集。对此数据集进行分类,就相当于对原始的失效数据进行分类。比如对此数据分为  $K$  类,将这  $K$  个类别与  $K$  种可靠性模型一一对应的,便可以得到可靠性模型与失效数据集之间的对应关系<sup>[1]</sup>。

#### 1.2 Kohonen 网络的原理<sup>[3]</sup>

1981 年 T. Kohonen 提出了自组织特征映射 SOFM 的概念,并给出了相应的 Kohonen 神经网络的模型。由于 Kohonen 网络的自调整过程模仿了人类大脑的自组织过程,因而它可以用来模仿有关外界信息在人脑中自组织地形成概念的过程。Kohonen 网络可以作为很好的聚类工具,与其他可分类的 BP 或 RBF 网络相比,它不需要对大量的已有数据进行训练,属于无监督的学习方式。Kohonen 网络能够快速的把输入按照其特征进行聚类,且分类的准确性较高。由于用来训练网络的数据很少,所以本文采用了 Kohonen 网络这种无监督的学习方法。Kohonen 网络可以作为一种样本特征检测器,在样本排序、样本分类以及样本检测方面有广泛的应用。

##### 1.2.1 Kohonen 网络模型<sup>[4]</sup>

Kohonen 网络使输入样本通过竞争学习后,功能相同的输入靠得比较近,不同的分得比较开,以此将一些无规则的输入自动排开,在联结权的调整过程中,使权的分布与输入样本的概率密度分布相似。

图 1 给出了由输入层和竞争层组成的无隐层 Kohonen 网

收稿日期:2005-04-12;修订日期:2005-07-02 基金项目:国防基础研究基金项目

**作者简介:**吴勤(1981-),男,江苏苏州人,硕士研究生,主要研究方向:复杂武器系统可靠性分析;侯朝桢(1938-),男,四川自贡人,教授,博士生导师,主要研究方向:分布式计算机控制、智能控制和容错控制、网络系统的可靠性和容错技术;原菊梅(1965-),女,山西河津人,讲师,博士研究生,主要研究方向:网络控制。

络模型。输入层是由  $N$  个神经元组成的,一维序列,输入层结点数对应于输入模式空间的维数;竞争层是由  $M \times N = H$  个神经元组成的二维平面阵列,竞争层也就是输出层,它的结点数对应于映射后的模式空间维数。可以是一维或二维点阵。输出层结点与输入层结点为全连接,输入层、输出层内部结点间没有连接关系。输出层内的每个神经元与其邻域连接,此连接是相互激励的关系,训练后输出层不同结点代表不同的分类模式,所以 SOFM 的输出层也称特征映射层。

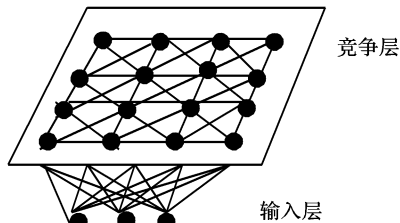


图1 Kohonen网络结构

### 1.2.2 网络训练过程

自组织映射神经网络(SOFM)这种自组织聚类过程是系统在自主、无导师指导的条件下完成的。Kohonen 网络的学习过程可分为以下两步:

第一步,神经元竞争学习过程。对于每一个输入向量,通过输入向量值与权重值之间的比较,在神经元之间产生竞争,权重向量与输入模式最相近的神经元被认为对于输入模式反映最强烈,将其标定为获胜的神经元,并称此神经元为输入模式的像,相同的输入向量会在输出层产生相同的像。

第二步,竞争层神经元的侧反馈过程。这种神经元之间的局部影响作用遵从以下规则:1)以获胜神经元为圆心,对临近的神经元表现为兴奋性侧反馈。2)以获胜神经元为圆心,对远邻的神经元表现为抑制性侧反馈。几何上相近的输出节点所连接的权重向量既有联系(即相似性),又相互有区别,保证了对于某一类输入模式,获胜节点能作出最大的响应,而相邻节点作出较少响应。

### 1.2.3 学习算法

1) 为网络中每一个神经元的权重赋较小的随机值;

2) 设网络输入向量  $\mathbf{X} = [X_0, X_1, \dots, X_{n-1}]$ , 神经元  $I$  的权重向量  $\mathbf{W}_i = [W_0, W_1, \dots, W_{n-1}]$ , 求得  $I_i = \sum_{j=0}^{N-1} \|x_j - w_{ij}\|$ ,  $\forall i$ ;

3) 令  $I_c = \min\{I_i\}$ , 以  $c$  所对应的神经元为竞争获胜的神经元;

4) 在以  $c$  为中心,  $N_c$  范围内的神经元, 按照公式  $w_{ij}(t+1) = w_{ij}(t) + a[x_j - w_{ij}(t)]$  调节权重,  $N_c$  范围外的神经元权重不调整, 其中,  $a$  是学习增益系数;

5) 取下一个输入向量, 回到 2) 反复执行。

## 2 基于 Kohonen 网络的模型选择过程

给定的软件失效数据, 无法直接用 Kohonen 网络进行分类, 必须首先对给定的数据进行编码。经编码后的失效数据作为 Kohonen 网络的输入, 通过 Kohonen 网络进行聚类, 就可以实现对软件可靠性模型进行选择。

### 2.1 软件可靠性模型

在本文的方法中, 提出 3 个软件可靠性模型<sup>[5,6]</sup>用作初选模型, 分别为 JM、GO、LV。这三种模型是目前在可靠性分析中应用比较广泛的三种模型。其中 JM 模型是最具代表性的

早期软件可靠性马尔可夫过程的数学模型; 而 GO 模型是最著名的 NHPP 过程模型; LV 模型是典型的 Bayesian 模型。这三种模型既包括随机过程模型又包括了非随机过程模型。能用来分析大部分软件的可靠性。给定的软件失效数据经编码后将分为 3 类, 每类与这 3 种模型相对应。

### 2.2 编码方法<sup>[1]</sup>

许多研究人员为了比较软件可靠性模型的应用能力, 提出了许多用于评价的准则。选用比较成熟的 5 种评价准则: 模型拟合性、模型预计有效性(常用序列似然度量)、模型偏差、模型偏差趋势和模型噪声, 评价准则的定义与及其具体计算方法可参见文献[5]。同时给三种模型编号, JM 为 1, GO 为 2, LV 为 3。

编码步骤:

1) 对给定的软件失效数据, 分别用 3 种模型计算第 1 种评价准则(模型拟和性)的值。并排出优劣顺序, 对模型拟和性适应性最好的模型打 3 分, 次好的为 2 分, 最差的为 1 分。分数最高的模型标号为  $x_1$ 。

2) 同 1) 相同的方法分别计算剩余 4 种评价准则得到  $x_2, x_3, x_4, x_5$ 。

3) 向量  $\{x_1, x_2, x_3, x_4, x_5\}$  就是某一软件失效数据对应的编码。

如表 1 给出的是文献[5]中 JPL 项目的数据组 J2, 括号内的是打分的结果, 其编码后的结果为  $\{1, 3, 3, 3, 2\}$ 。

表 1 数据组 J2 的编码

准则	JM	GO	LV	适应性最优 模型编号
模型拟和性	0.162(3分)	0.182(1分)	0.171(2分)	1
模型精确度	1074(2分)	1075(1分)	1051(3分)	3
模型偏差	0.3378(2分)	0.3378(2分)	-0.2592(3分)	3
模型偏差趋势	0.4952(2分)	0.4954(1分)	0.1082(3分)	3
模型噪声	2.607(2分)	2.593(3分)	23.33(1分)	2

### 2.3 模型选择过程

为了使通过 Kohonen 网络聚类分析的每个类别与具体的模型相对应, 采用基准数据的办法。所谓基准数据指此数据对某可靠性模型在模型拟合性、序列似然度、模型偏差、偏差趋势及模型噪声各个方面的适应性均为最佳的数据。在这里因为选用了三个模型作初选模型, 所以有三个基准数据, 其编码后的结果应为:  $\{1, 1, 1, 1, 1\}$  (适合 JM 模型);  $\{2, 2, 2, 2, 2\}$  (适合 GO 模型);  $\{3, 3, 3, 3, 3\}$  (适合 LV 模型);

每个软件的失效数据经编码后成为一个五维向量。如果两个向量划分在同一个类别中, 称为等价向量。将生成向量的对应失效数据称为等价失效数据。等价的失效数据可以选用同一种软件可靠性模型。所以分类的结果中与基准数据分为同一组的数据所适合的可靠性模型也应该和基准数据的相同。在本方法中, 把给定的一系列失效数据经过编码后作为 Kohonen 网络的输入, 指定输出分为三类, 这样通过网络自动地把失效数据分为三类。因此, 对于给定的某软件失效数据通过查看与何种基准数据聚为同一类就可以判断该软件适合何种可靠性模型。

## 3 仿真分析

在这种方法中, 使用了 20 组软件可靠性失效数据, 一部分来源于贝尔实验室 John D. Musa 在文献[7]上公布的软件失效数据。剩下的来自于丹麦信息与数学建模技术大学提供

的教学数据包。这 20 组数据记为 D1, D2, ..., D20。其中 D1, D2, D3 是定义的基准数据, 编码为  $D1 = \{1, 1, 1, 1, 1\}$ ,  $D2 = \{2, 2, 2, 2, 2\}$ ,  $D3 = \{3, 3, 3, 3, 3\}$  首先将给定的原始数据编码, 编码后用 Kohonen 网络进行分析。为了说明聚类效果, 表 2 给出了本文的方法与工程上应用成熟的累加排序方法对这 20 组数据分析的结果对比, 只有 D7 与 D17 出现不一致, 其错误率为 10%, 说明该方法是可行的。在实际使用的过程中可以结合多种选择模型的方法综合得出结论。本仿真

采用了 Matlab 中的神经网络工具箱, 通过运算发现, 当训练 70 次左右的时候, 其准确性最好。为了比较算法的优劣, 除了利用 Kohonen 网络, 还与目前对可靠性模型选择研究的高斯混合模型方法<sup>[1]</sup> 及比较误差系数法<sup>[9,10]</sup> 两种主要方法相比较。同样用上述 20 组数据与工程上应用成熟的累加排序方法进行比较。采用高斯混合模型方法计算错误率约为 15%, 采用比较误差系数方法的错误率为 23%。证明了此方法的有效性 & 准确性。

表 2 实验数据分类结果

编号	编码结果	累加排序结果	Kohonen 网络分类结果	编号	编码结果	累加排序结果	Kohonen 网络分类结果
D1	1 1 1 1 1	JM(1)	JM(1)	D11	1 3 3 3 3	YO(3)	YO(3)
D2	2 2 2 2 2	GO(2)	GO(2)	D12	2 1 2 1 1	JM(1)	JM(1)
D3	3 3 3 3 3	YO(3)	YO(3)	D13	3 3 3 2 2	YO(3)	YO(3)
D4	3 3 3 1 2	YO(3)	YO(3)	D14	2 2 3 1 1	GO(2)	GO(2)
D5	2 1 1 3 1	JM(1)	JM(1)	D15	1 2 3 1 2	GO(2)	GO(2)
D6	2 1 1 3 1	JM(1)	JM(1)	D16	1 2 1 3 1	JM(1)	JM(1)
D7	3 1 2 1 3	JM(1)	GO(2)	D17	2 1 2 1 3	GO(2)	JM(1)
D8	2 3 1 1 1	JM(1)	JM(1)	D18	2 3 3 3 3	YO(3)	YO(3)
D9	2 2 2 2 1	GO(2)	GO(2)	D19	1 1 1 2 3	JM(1)	JM(1)
D10	3 3 1 3 1	YO(3)	YO(3)	D20	1 3 3 3 2	YO(3)	YO(3)

#### 4 结语

关于软件可靠性模型选择问题的研究已经有了一定的发展。本文提出的基于 Kohonen 网络的模型选择方法, 是综合利用多种模型评价准则进行模型选择的一种尝试。该方法具有结构简单, 计算速度快和计算精度高等特点, 通过仿真试验证明了该方法的有效性和可行性, 可适用于软件可靠性模型的选择, 从而为软件模型的选择提供了一条新的思路。

#### 参考文献:

- [1] 汪浩, 刘超, 金茂忠. 基于聚类思想的软件可靠性模型选择[J]. 计算机工程与应用, 2002, 21(4): 21-24.
- [2] ELEWA SA. Development of an Environment for Software Reliability Model Selection[D]. Doctoral thesis. School of Engineering of the Air Force institute of Technology Air University, ADA256609, 1992: 225.
- [3] 丛爽. 神经网络、模糊系统及其在运动控制中的应用[M]. 合肥: 中国科学技术大学出版社, 2001.
- [4] 白瑞祥, 惠鸿忠, 宋辉. 基于自组织映射神经网络的聚类分析系统研究[J]. 化工自动化及仪表, 2004, 31(5): 29-31.
- [5] LYU MR. Handbook of software reliability Engineering [M]. McGraw-Hill publishing, 1995.

- [6] 徐仁佐. 软件可靠性模型与应用[J]. 质量与可靠性, 1994, 18(3): 27-31.
- [7] MUSSA JD. Software Reliability Data. DACS, RADCI DB/OL. <http://www.dacs.dtic.mil/databases/sled/swrel.shtml>, 1980.
- [8] Florin POPENTIU VLADICESCU. Performance Evaluation of Computers Courses. Technical University of Denmark Informatics and Mathematical Modelling[EB/OL]. <http://www.imm.dtu.dk/popentiu/pec/pec.html>, 2005.
- [9] 宋晓秋. 软件可靠性及维修性评估工具[J]. 电子产品可靠性与环境试验, 1999, 24(3): 16-21.
- [10] 黄智伟. 软件测试和可靠性评估模型选择[J]. 电子计算机, 1999, 4(6): 24-26.
- [11] 徐仁佐, 袁凌, 陈波. 软件可靠性模型应用中的不一致性与软件可靠性专家系统[J]. 计算机应用与软件, 2001, 21(4): 36-41.
- [12] 田涛, 张凤鸣, 王昕. 一种基于模糊综合评判的软件可靠性模型选择方法[J]. 空军工程大学学报, 2002, 3(2): 56-59.
- [13] 邹丰忠, 徐仁佐. 软件可靠性多模型综合评估[J]. 同济大学学报, 2002, 30(10): 1183-1185.
- [14] 魏长春, 张净敏. 趋向性分析在软件可靠性模型中的应用[J]. 质量与可靠性, 2004, 18(6): 38-41.

(上接第 2330 页)

(如 CHI 值法) 有更高的聚类精度, 因而本文提出的方法较现有同类方法具有更大的优越性。本法较适用于文本数较多, 互信息较充分的场合, 是一种较为实用的文本聚类方法。

#### 参考文献:

- [1] PAL SK, et al. Web mining in soft computing framework: relevance, state of the art and future directions [J]. IEEE Transactions on Neural Networks, 2002, 13(5): 1163-1177.
- [2] KASKI S. Dimensionality Reduction by Random Mapping: Fast similarity Computation for Clustering[A]. Proceedings of international Joint Conference on Neural Networks(IJCNN98)[C]. IEEE Service Center, Piscataway, NJ, 1998. 413-418.
- [3] KURIMO M. Fast latent semantic indexing of spoken documents by using self-organizing maps[A]. IEEE international Conference on A-

coustics, Speech and signal Processing(ICASSP00)[C]. Istanbul, Turkey, 2000. 2425-2428.

- [4] 钱晓东, 王正欧. 基于 SOM 网络的随机映射文本降维方法[J]. 计算机应用, 2004, 24(5): 56-59.
- [5] DAGAN I, MARCUS S. Contextual word similarity and estimation from sparse data[J]. Computer Speech and Language, 1995, 9: 123-152.
- [6] SAHAMI M. Using machine learning to information access[D]. A Dissertation of Standford university, 1998.
- [7] KOHONEN T. self-organizing maps[M]. Springer, Berlin, 1995.
- [8] 王明春, 王正欧, 张楷, 等. 一种基于 CHI 值特征选取的粗糙集文本分类规则抽取方法[J]. 计算机应用, 2005, 25(5): 1026-1028.