【兵器与装备】

基于主成分分析法的网络系统生存能力评估*

崔　鹏，张耀坤，徐志伟

（军械工程学院，石家庄　050003）

摘要：运用主成分分析法对网络系统生存能力的多项指标进行综合聚集，在理想点的基础上建立综合优化决策模型，通过实例分析，表明应用该方法对网络系统生存能力进行评估是有效和可行的．

关键词：主成分分析法；优化决策模型；网络系统生存能力

中图分类号：TN711　　　**文献标识码**：A　　　**文章编号**：1006－0707(2008)06－0064－03

可生存性是指在遭受攻击、故障或意外事故时，系统能够及时地完成其关键任务的能力[1]．可生存性的中心思想是即使在遭受入侵后，或系统的重要部分遭受到损害或摧毁时，系统仍然能够完成任务，并具有在一定时间内修复被损坏服务的能力．随着信息技术的快速发展，网络攻击手段层出不穷，网络中的任何一个节点都可能因为攻击、故障和意外事故等而失效，为了保障网络系统在面临攻击的时候能够稳定、可靠地提供服务，必须考虑网络系统自身的生存能力．通过对网络系统生存能力的量化评估有助于对系统的生存性进行准确的评价和对比，从而进行合理的维护和技术改造，并进一步实现可生存性系统的分析和设计[2]．

1　网络系统生存能力评估体系[3]

可生存系统的关键特征是在面对攻击、故障、意外事故时完成基本服务的能力．为了较好地完成基本服务的能力，在遵循网络系统生存能力指标体系建立原则的基础之上，经广泛征求专家意见，以网络系统生存能力为目标层，从抵抗攻击、识别攻击、服务恢复和自适应演化 4 个方面建立评估指标体系，如表 1 所示．

2　主成分分析法

主成分分析法是将多个指标化为少数指标的一种统计方法．利用主成分分析法对多维变量进行降维，降维后的变量是原变量的线性组合，并能反映原变量绝大部分的信息，使信息的损失最小，对原变量的综合解释能力强．该方法通过主成分的方差贡献率来表示变量的作用，可避免在系统分析中对权重的主观判断，使权重的分配更合理，尽可能地减少重叠信息的不良影响，克服变量之间的多重相关性，使系统分析简化．当主成分变量所包含的指标信息量占原始指标信息量的 85% 以上时，认为分析达到了效果．

表 1　网络生存能力评估体系

一级指标	二级指标	三级指标
网络生存能力（A）	抵抗攻击（B_1）	攻击结果（C_1）
		危险程度值（C_2）
		关键步骤成功率（C_3）
	识别攻击（B_2）	识别率（C_4）
		识别时间（C_5）
	服务恢复（B_3）	恢复时间（C_6）
		恢复程度（C_7）
		服务的关键性（C_8）
		服务恢复率（C_9）
	自适应演化（B_4）	演化延迟（C_{10}）
		自适应周期（C_{11}）

主成分的求解步骤为：

步骤 1　为了消除各指标之间因为度量单位不同引起的差异，将数据 $x_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 采用 Z-Score 法进行标准化处理：

$$Z_{ij}=(x_{ij}-\overline{x_j})/\sigma_j \tag{1}$$

其中，$\overline{x_j}=\frac{1}{m}\sum_{i=1}^{m}x_{ij}$；$\sigma_j=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_{ij}-x_j)^2}$．得到标准化样本决策矩阵 $Y=(y_{ij})_{m\times n}$．

步骤 2　计算所有样本的指标相关矩阵 $R=(r_{ik})_{m\times n}$，

其中，$r_{jk}=\frac{1}{m-1}\sum_{i=1}^{m}y_{ij}\cdot y_{ik}(i,j=1,2,\cdots,n)$　　(2)

步骤 3　求相关矩阵 R 的特征值 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n\geqslant 0$

* **收稿日期**：2008－06－23

作者简介：崔鹏（1982—），男，辽宁葫芦岛人，硕士研究生，主要从事军事运筹学研究．

及相应的特征向量

$$\alpha_j=[\alpha_{j1},\alpha_{j2},\cdots,\alpha_{jn}]^{\mathrm{T}}(j=1,2,\cdots,n) \tag{3}$$

步骤 4　通过 λ_i 计算各分量的方差贡献率:

$$p_i=\lambda_i/\sum_{i=1}^{n}\lambda_i(i=1,2,\cdots,n) \tag{4}$$

选取满足 $\sum_{i=1}^{p}\lambda_i/\sum_{j=1}^{n}\geqslant e$(一般取 $e=0.85$)的前 p 个主成分作为新的决策指标,从而得到低维指标的主成分决策矩阵 $Z=(z_{ij})_{m\times p}=[z_1,z_2,\cdots,z_p]$

根据主成分分析的结果,可得主成分指标权重

$$\mu_j=\lambda_j/\sum_{k=1}^{p}\lambda_k(j=1,2,\cdots,p) \tag{5}$$

进而可构造主成分加权决策矩阵 $U=(u_{ij})_{m\times p}=[\mu_1z_1,\mu_2z_2,\cdots,\mu_pz_p]$.

3　建立优化决策模型[4]

定义 1　对主成分加权决策矩阵 $U=(u_{ij})_{m\times p}$的指标属性值,取 $u_j^+=\max\{u_{ij}|i=1,2,\cdots,m\}$,$u_j^-=\min\{u_{ij}|i=1,2,\cdots,m\}(j=1,2,\cdots,p)$,则称由 $A^+=(u_1^+,u_2^+,\cdots,u_p^+)$和 $A^-=(u_1^-,u_2^-,\cdots,u_p^-)$构成的方案分别为主成分正理想方案(正理想点)和主成分负理想方案(负理想点).

定义 2　称

$$d_i^+=[\sum_{j=1}^{p}u_{ij}-u_j^+]^{\frac{1}{2}}\quad(i=1,2,\cdots,m) \tag{6}$$

$$d_i^-=[\sum_{j=1}^{p}u_{ij}-u_j^-]^{\frac{1}{2}}\quad(i=1,2,\cdots,m) \tag{7}$$

为方案 $A_i(i=1,2,\cdots,m)$对主成分正理想方案和主成分负理想方案的偏离度.

定义 1 中的主成分正理想方案和主成分负理想方案显然不存在,否则就无需决策,作此定义的宗旨是把它们作为衡量其他可行方案的准绳以权衡各方案的优劣.假设综合主成分指标 q_i 为方案A_i 期望排序,则 $q_id_i^+$ 为方案 A_i 距离主成分正理想方案的期望偏离度,$(1-q_i)d_i^-$ 为方案 A_i 距离主有必要负理想方案的期望偏离度.我们的目标是,使每一方案 A_i 要么偏离主成分负理想方案(贴近主成分正理想方案),要么偏离主成分正理想方案(贴近主成分负理想方案),这样就可以将各方案的优劣性明显地区别开来,为此建立如下综合优化决策模型:

$$\begin{aligned}&\max F(q_1,q_2,\cdots,q_m)=\\&\max\{\sum_{i=1}^{m}(q_id_i^+)^2+[(1-q_i)d_i^-]^2\}=\\&\max\sum_{i=1}^{m}[q_i^2\sum_{i=1}^{m}(u_{ij}-u_j^+)^2+(1-q_i)^2\sum_{j=1}^{p}(u_{ij}-u_j^-)^2]\end{aligned} \tag{8}$$

令 $\frac{\partial F}{\partial q_i}=0$,得

$$q_i=\frac{\sum_{j=1}^{p}(u_{ij}-u_j^-)^2}{\sum_{j=1}^{p}(u_{ij}-u_j^+)^2+\sum_{j=1}^{p}(u_{ij}-u_j^-)^2} \tag{9}$$

由式(9)可得到在主成分分析和理想点基础上的各方案总体优化决策排序向量 $q=(q_1,q_2,\cdots,q_m)$,决策者可据此结果进行决策.

4　实例评估

以某 4 个单位指挥所网络系统为例,采用专家咨询法对以上指标进行了预测,结果如表 2 所示.现通过主成分分析法对各单位的网络系统生存能力进行评估.

表 2　各单位评价指标预测表

指标	单位			
	A_1	A_2	A_3	A_4
C_1	降级服务	普通服务	升级服务	普通服务
C_2	高	中	低	中
C_3	15%	35%	50%	40%
C_4	78%	85%	90%	80%
C_5	0.5	0.4	0.3	0.4
C_6	0.8	1.0	1.2	0.9
C_7	好	中	差	中
C_8	高	高	中	中
C_9	100%	95%	80%	85%
C_{10}	0.4	0.2	0.4	0.5
C_{11}	0.5	0.4	0.5	0.2

分析:对表中各目标的定性指标进行量化,得到原始指标数据矩阵 X 为

$$X=\begin{pmatrix}1&1&0.15&0.78&0.5&0.8&5&5&1.00&2.5&2.0\\3&3&0.35&0.85&0.4&1.0&3&5&0.95&5.0&2.5\\5&5&0.50&0.90&0.3&1.2&1&3&0.80&2.5&2.0\\3&3&0.40&0.80&0.4&0.9&3&3&0.85&2.0&5.0\end{pmatrix}$$

再由式(1)得标准化样本决策矩阵

$$Y=\begin{pmatrix}-1.2247&-1.2247&-1.3587&-0.9763&1.2247&-1.0247&1.2247&0.8660&1.0954&-0.3693&-0.6093\\0&0&0&0.3254&0&0.1464&0&0.8660&0.5477&1.4771&-0.2611\\1.2247&1.2247&1.0190&1.2552&-1.2247&1.3175&-1.2247&-0.8660&-1.0954&-0.3693&-0.6093\\0&0&0.3397&-0.6044&0&-0.4392&0&-0.8660&-0.5477&-0.7385&1.4797\end{pmatrix}$$

再由式(2)得样本的指标相关矩阵 R,并可计算出 R 的特征值和贡献率如表 2 所示.

$$
R=\begin{pmatrix}
1.0000 & 1.0000 & 0.9707 & 0.9110 & -1.0000 & 0.9562 & -1.0000 & -0.7071 & -0.8944 & 0 & 0\\
1.0000 & 1.0000 & 0.9707 & 0.9110 & -1.0000 & 0.9562 & -1.0000 & -0.7071 & -0.8944 & 0 & 0\\
0.9707 & 0.9707 & 1.0000 & 0.8001 & -0.9707 & 0.8619 & -0.9707 & -0.7845 & -0.9303 & -0.0418 & 0.2365\\
0.9110 & 0.9110 & 0.8001 & 1.0000 & -0.9110 & 0.9891 & -0.9110 & -0.3758 & -0.6451 & 0.2747 & -0.3831\\
-1.0000 & -1.0000 & -0.9707 & -0.9110 & 1.0000 & -0.9562 & 1.0000 & 0.7071 & 0.8944 & 0 & 0\\
0.9562 & 0.9562 & 0.8619 & 0.9891 & -0.9562 & 1.0000 & -0.9562 & -0.5071 & -0.7483 & 0.1441 & -0.2888\\
-1.0000 & -1.0000 & -0.9707 & -0.9110 & 1.0000 & -0.9562 & 1.0000 & 0.7071 & 0.8944 & 0 & 0\\
-0.7071 & -0.7071 & -0.7845 & -0.3758 & 0.7071 & -0.5071 & 0.7071 & 1.0000 & 0.9487 & 0.6396 & -0.5025\\
-0.8944 & -0.8944 & -0.9303 & -0.6451 & 0.8944 & -0.7483 & 0.8944 & 0.9487 & 1.0000 & 0.4045 & -0.3178\\
0 & 0 & -0.0418 & 0.2747 & 0 & 0.1441 & 0 & 0.6396 & 0.4045 & 1.0000 & -0.3428\\
0 & 0 & 0.2365 & -0.3831 & 0 & -0.2888 & 0 & -0.5025 & -0.3178 & -0.3428 & 1.0000
\end{pmatrix}
$$

表 2　R 的特征值和贡献率

特征值	λ_1	λ_2	λ_3	$\lambda_4\cdots\lambda_8$
	8.0245	2.2498	0.7257	0
贡献率	0.7295	0.2045	0.066	0
累积贡献率	0.7295	0.9340	1	1

由表 2 可得 $\sum_{i=1}^{2}\lambda_i/\sum_{j=1}^{11}\lambda_j=93.40\%>85\%$，因此取前两项主成分即可达到决策分析的要求. λ_1,λ_2 所对应的特征向量为

$$a_1=[0.3519\quad 0.3519\quad 0.3445\quad 0.3108\quad -0.3519\quad 0.3306\quad -0.3519\quad -0.2685\quad -0.3272\quad -0.0272\quad 0.0164]^{\mathrm T}$$

$$a_2=[-0.0503\quad -0.0503\quad 0.0640\quad -0.3126\quad 0.0503\quad -0.2236\quad 0.0503\quad -0.4279\quad -0.2481\quad -0.5453\quad 0.5453]^{\mathrm T}$$

选取前两个主成分指标，$Z_j=Y\cdot a_j(j=1,2)$，从而得主成分决策矩阵 $Z=(z_{ij})_{4\times2}=[Z_1,Z_2]$，再由式(4)算出主成分指标权重，$u_1=0.7295$，$u_2=0.2045$. 进而可以构造主成分加权决策矩阵 $U=(u_{ij})_{4\times2}=[u_1Z_1,u_2Z_2]_{4\times2}$，并由此得到主成分正、负理想方案分别为：

$$\begin{cases}A^+=(2.5471\quad 0.4138)\\ A^-=(-2.4986\quad -0.3248)\end{cases}$$

由优化决策模型(8)及式(9)，得到各单位总体优化决策排序值如表 3 所示.

表 3　各单位网络系统生存能力排序

目标	m_1	m_2	m_3	m_4
排序值 q_i	0.0037	0.3863	0.9908	0.5776
排序	4	3	1	2

由表 3 得出，网络系统生存能力按照由大到小的顺序依次为 $m_3>m_4>m_2>m_1$. 从评估结果来看，其结果较为合理、客观，且符合实际情况.

5　结束语

主成分分析法是一种科学、简便的评估方法. 采用主成分分析法，能比较客观地对网络系统生存能力进行评估，其准确率较高，原理也比较简单，所得的结论比较客观；并且指标增加时，这种方法在处理数据、进行有效决策方面更具优势，可以很快得到科学的结果，为有效评估网络系统生存能力提供科学依据.

参考文献：

[1] Westmark R. A Definition for Information System Survivability[C]//Proceeding of the 37th Hawaii Internal Conference on System Sciences(HICSS'04). [S.l.]: IEEE, 2004.

[2] Somesh Jha. Survivability Analysis of Networks Systems [EB/OL]. [2008-07-16]. http://www.cs.wisc.edu/~jha/jha-paper/security/icse-2001.pdf.

[3] 王健，王慧强，赵国生. 基于不确定型 AHP 的网络生存能力模糊综合评估[J]. 计算机科学，2006: 33(6): 73.

[4] 刘家学，陈世国. 基于主成分分析的投资决策[J]. 运筹与管理，2006(2): 77.