

基于人工神经网络的系统可靠度预测

樊立天,江金达,夏景涛,崔飞易,缪吉昌,王婷婷,夏红林,王胜军,陈宏文
南方医科大学南方医院医学工程科,广东 广州 510515

【摘要】提出一种基于人工神经网络的可靠度预测方法,以系统故障时间和对应的中位秩估计值训练网络,在系统故障时间范围内选取大量的故障时间点并求其预测的累积分布函数值,然后结合样条回归法求出系统累积分布函数曲线,概率密度函数曲线及故障率函数曲线。为验证人工神经网络模型的优越性,以婴儿培养箱等5个系统的故障数据为例,用决定系数 R^2 、均方误差和对数似然函数,与Weibull、Fréchet、Logistic等统计分布模型进行对比,结果表明人工神经网络拟合效果最优。

【关键词】人工神经网络;统计分布;可靠度预测;拟合精度

【中图分类号】R318;TB114.3

【文献标志码】A

【文章编号】1005-202X(2023)02-0232-06

System reliability prediction based on artificial neural network

FAN Litian, JIANG Jinda, XIA Jingtao, CUI Feiyi, MIAO Jichang, WANG Tingting, XIA Honglin, WANG Shengjun, CHEN Hongwen
Department of Medical Engineering, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China

Abstract: A system reliability prediction method based on artificial neural network is proposed. The network is trained with the system failure data and the corresponding estimate of median rank estimate. A large number of failure time points are selected in the system failure duration, and their predicted cumulative distribution function values are found. Subsequently, the cumulative distribution function curve, probability density function curve, and failure rate function curve of the system are derived using spline regression method. To verify the superiority of the artificial neural network model, taking the failure data of 5 systems such as infant incubator as an example, the proposed model is compared with the statistical distribution models such as Weibull, Fréchet and Logistic using the coefficient of determination R^2 , mean square error and log-likelihood function. The results show that the artificial neural network fits the best.

Keywords: artificial neural network; statistical distribution; reliability prediction; fitting accuracy

前言

随着系统集成度越来越高,系统的可靠性分析面临着愈加严重的挑战。系统退化水平预测是可靠性分析的基本问题,对于制定保养计划具有重要参考意义^[1]。传统系统退化水平预测大多采用基于概率统计的方法,典型的概率统计分布模型有Normal分布^[2]、Exponential分布^[3]、Weibull分布^[4]、Logistic分布^[5]、Fréchet分布^[6]等,通过确定分布函数的特征值及失效概率,在指定可靠度下预测设备退化水平。但这些方法的前提要积累大量的设备失效数据,而

设备失效数据的来源大致可分为两类:一类是设备上市前在实验室进行的模拟或者加速实验;另一类是上市后投入复杂的真实世界环境中发生的失效或故障。这两种获取数据的方法都源自系统本身,对于成本低廉的系统可以采用统计学方法,但若研究造价高昂的系统可靠性,传统的统计学方法大样本的条件难以满足。此外,传统统计分布模型的故障率函数较为单一,难以满足复杂的真实世界环境。比如,Exponential分布的前提是假设系统故障率恒定,忽略了复杂的外部环境影响^[7];Weibull分布故障率函数只能是恒定或单调的,无法一次性描述复杂的“浴盆曲线”或“单峰曲线”^[8];即便是在可靠性工程中广泛使用的“浴盆曲线”,也只能描述10%~15%故障情况^[9]。用于系统退化水平预测的传统工具越来越不能满足行业对精度和准确度日益增长的需求,然而基于机器学习很适合解决以上难题。

机器学习已渗透到众多学科领域,可靠性和安

【收稿日期】2022-10-09

【基金项目】国家重点研发计划(2019YFC0121908);广东省科技计划项目(2017ZC0068);广东省自然科学基金(2017ZC0069)

【作者简介】樊立天,研究方向:机器学习, E-mail: fanli960@163.com

【通信作者】陈宏文,高级工程师,研究方向:医疗设备质控管理, E-mail: Chw47922@126.com

全工程也不例外。机器学习是一种从数据中学习和揭示规律的方法,又可用于推理和预测^[10]。文献[11]研究了基于人工神经网络(Artificial Neural Network, ANN)和模糊逻辑分类器的专家系统,用实时测量的温度误差等数据来预测婴儿培养箱性能。文献[12]开发一种用于除颤器硬件性能预测的自动化系统,分别使用5种不同的机器学习算法对具有不同属性数量的数据集进行评估,发现随机森林分类器准确率最高。文献[13]通过ANN从加速老化数据中预测核电站电缆的寿命和可靠性。本文提出一种基于ANN的系统退化水平预测方法,并将其应用在5组故障数据集上,结果发现该方法与传统统计分布模型相比具有更高的灵活性和拟合精度。

1 ANN模型

ANN可以表示为节点的连接层,每个节点接收几个加权的输入,并产生一个输出。一个常用的激活函数是ReLU(Rectified Linear Unit)函数:

$$g(x) = \max(0, x) \quad (1)$$

图1中的圆圈代表节点,也称为感知器,激活函数位于节点中,取输入的 x 值并给予其权重 w ,将其相加并插入激活函数 g 中,激活函数的输出用 h 表示。激活函数 $h_1^{(2)}$ 对应于第2个隐藏层的第一函数,可表示为:

$$h_1^{(2)} = g(x_1 w_1 + x_2 w_2 + \dots + x_n w_n + b_1^{(2)}) \quad (2)$$

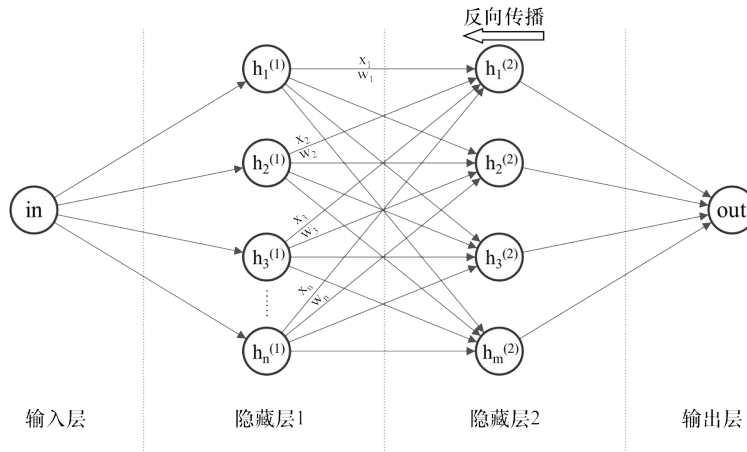


图1 ANN示意图

Figure 1 ANN diagram

ANN模型将收到一个数字向量,即直到失效的时间,并将输出另一个代表概率的向量。训练时首先用随机参数启动程序,然后通过比较与预期值的输出,再调整参数,使输出更接近于预期值,不断重复的比较、调整,直到误差低于预期值。本研究使用均方误差(Mean Squared Error, MSE)作为损失函数,因其梯度值是动态变化的,在回归问题中可以快速收敛,节省训练时间。

2 拟合效果评价指标

本研究使用决定系数 R^2 、MSE、贝叶斯信息准则(Bayesian Information Criterion, BIC)和赤池信息准则(Akaike Information Criterion, AIC)来比较模型的拟合精度,公式如下:

$$\hat{F}_i = \frac{i - 0.3}{n + 0.4} \quad (3)$$

$$L = \sum_{i=1}^n \ln f_{\text{pre}} \quad (4)$$

$$\text{BIC} = k \ln(n) - 2 \ln L \quad (5)$$

$$\text{AIC} = 2k - 2 \ln L \quad (6)$$

$$\text{AIC}_c = \text{AIC} + \frac{2k(k+1)}{n-k-1} \quad (7)$$

$$\text{MSE} = \frac{\sum_{i=1}^n [F_{\text{pre}} - \hat{F}]^2}{n} \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n [\hat{F}_i - \hat{F}_{\text{pre}}]^2}{\sum_{i=1}^n [\hat{F}_i - \bar{\hat{F}}]^2} \quad (9)$$

其中, i 为故障次序, n 为故障数据量, k 为模型参数数量,

L 为最大似然函数, $\bar{\hat{F}}_i = \frac{\sum \hat{F}_i}{n}$, f_{pre} 、 \hat{F}_{pre} 为预测的概率密度函数(Probability Density Function, PDF)、累积分布函数(Cumulative Distribution Function, CDF)估计值,当样本量较小时 AIC_c 比AIC更准确。 R^2 越趋近于1,MSE、 AIC_c 及BIC值越小,代表模型拟合效果越好。

为验证 ANN 模型的性能,本研究同时使用 Weibull、Exponential、Fréchet、Logistic、Normal 分布模型进行分析,其 PDF 和 CDF 表达式如表 1 所示, t 为故障时间。故障率函数(Faliure Rate Function, FRF)公式为:

FRF=PDF/(1-CDF)

(10)

表 1 Weibull、Exponential、Fréchet、Logistic 和 Normal 分布的 PDF 和 CDF
Table 1 PDF and CDF of Weibull, Exponential, Fréchet, Logistic and Normal distributions

模型	概率密度函数	累积分布函数
Weibull	$f(t; \alpha, \beta) = \frac{\beta}{\alpha^\beta} t^{(\beta-1)} \exp[-(\frac{t}{\alpha})^\beta]$	$F(t; \alpha, \beta) = 1 - \exp[-(\frac{t}{\alpha})^\beta]$
Exponential	$f(t; \sigma) = \frac{1}{\sigma} \exp(-\frac{t}{\sigma})$	$F(t; \sigma) = 1 - \exp(-\frac{t}{\sigma})$
Fréchet	$f(t; \mu, \sigma) = \exp[-\exp(-\frac{\log(t) - \mu}{\sigma})] \exp(-\frac{\log(t) - \mu}{\sigma}) \frac{1}{t\sigma}$	$F(t; \mu, \sigma) = \exp[-\exp(-\frac{\log(t) - \mu}{\sigma})]$
Logistic	$f(t; \mu, \sigma) = \frac{1}{\sigma} \phi_{\logis}(\frac{t - \mu}{\sigma})$	$F(t; \mu, \sigma) = \Phi_{\logis}(\frac{t - \mu}{\sigma})$
Normal	$f(t; \mu, \sigma) = \frac{1}{\sigma} \phi_{\text{nor}}(\frac{t - \mu}{\sigma})$	$F(t; \mu, \sigma) = \Phi_{\text{nor}}(\frac{t - \mu}{\sigma})$

μ :位置参数; σ :尺度参数;Weibull: $\alpha = \exp(\mu), \beta = 1/\sigma$;Logistic: $\phi_{\logis}(z) = \frac{\exp(z)}{[1 + \exp(z)]^2}, \Phi_{\logis}(z) = \frac{1}{1 + \exp(-z)}$;Normal: $\phi_{\text{nor}}(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}), \Phi_{\text{nor}}(z) = \int_{-\infty}^z \phi_{\text{nor}}(w) dw$

3 数据分析

本研究使用 5 组与可靠性相关的数据集(表 2)评估 ANN 在可靠度预测中的有效性。将这些故障数据按升序排序,然后使用中位秩估计法[式(3)]计算 CDF 估计值。为了更好地与统计分布模型进行比较,所有的数据都被用来训练 ANN 模型,故障时间作为输入,中位秩估计值作为输出,由 10 个感知器组成中间的隐藏层。为了更直观地看到 ANN 模型的 CDF,在故障时间范围内选取大量的测试时间点并求其 CDF 预测值,然后使用样条回归法拟合出平滑的 CDF 曲线。与统计分布模型相同,ANN 模型的 CDF 也应满足单调递增的条件,且返回值在[0, 1]之间,因此 PDF 必须要为非负值。图 2a 中的黑色圆圈为故

障时间点,红线为 ANN 模型求得的 CDF 曲线,图 2b 和图 2c 分别为单峰型的 PDF 和 FRF 曲线,其 $R^2=0.991\ 2$, $MSE=0.000\ 7$,说明 ANN 模型拟合精度非常高。

表 2 故障数据集
Table 2 Fault dataset

数据集	数量	描述
1	16	婴儿培养箱故障数据
2	36	文献[14]提供的发电机失效时间数据
3	20	文献[15]提供的电子元件使用寿命
4	24	文献[16]提供的机械部件的故障时间数据
5	50	文献[17]提供的随机数据

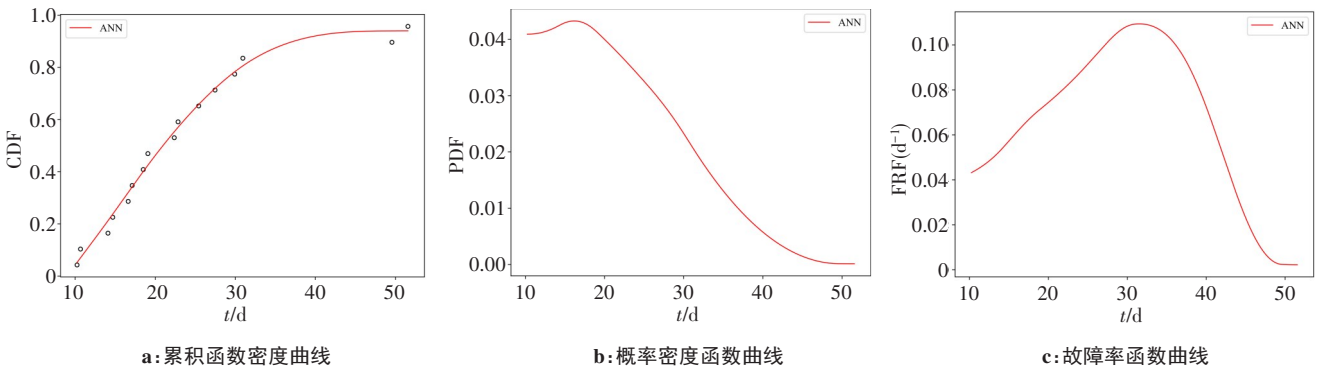


图 2 第 4 组数据集的 ANN 模型函数
Figure 2 Functions of the ANN model for the dataset 4

表3和表4给出了Weibull、Exponential、Fréchet、Logistic和Normal分布的位置参数和尺度参数估计值(本研究使用的Exponential分布为单参数分布,无位置参数)以及Weibull分布的 α 和 β 值,表5给出了评价模型拟合效果的 R^2 、 AIC_c 、BIC、MSE及 $-2\ln L$ 值和故障率函数形状。

表3 统计分布的位置参数和尺度参数

Table 3 Location and scale parameters of statistical distribution models

序号	数据集	模型	位置参数	尺度参数
1	1	Weibull	3.933 7	1.286 9
2	1	Fréchet	2.591 2	1.268 4
3	1	Normal	60.500 0	89.092 6
4	1	Logistic	42.449 6	37.384 0
5	1	Exponential	-	60.500 0
6	2	Weibull	0.838 0	1.226 1
7	2	Fréchet	-0.690 1	1.547 6
8	2	Normal	2.567 4	2.759 6
9	2	Logistic	2.153 9	1.433 3
10	2	Exponential	-	2.567 4
11	3	Weibull	0.712 5	0.836 1
12	3	Fréchet	-0.541 6	1.592 5
13	3	Normal	1.935 5	1.399 7
14	3	Logistic	1.843 7	0.804 4
15	3	Exponential	-	1.935 5
16	4	Weibull	3.297 4	0.462 2
17	4	Fréchet	2.842 8	0.400 6
18	4	Normal	23.822 5	11.782 9
19	4	Logistic	22.103 3	6.245 7
20	4	Exponential	-	23.822 5
21	5	Weibull	3.929 3	0.225 2
22	5	Fréchet	3.684 6	0.242 5
23	5	Normal	46.450 0	11.212 8
24	5	Logistic	46.013 5	6.507 0
25	5	Exponential	-	46.450 0

以 AIC_c 和BIC两项指标来看,Exponential分布性能更佳(图3a),但在第4、5组数据集中Exponential分布的 R^2 仅为0.666 7和0.333 5,这也说明传统分布拟合效果极不稳定,只对特定类型的数据效果较好。ANN模型的 R^2 值在5组数据集上均为最高,在第1、4、5组数据集上 $R^2>0.99$,说明拟合效果非常理想,除了第5组数据集,在其他数据集上都表现出非常明显

表4 Weibull分布的 α 和 β 值

Table 4 The α and β values of Weibull distribution

数据集	α	β
1	51.094 8	0.777 1
2	2.311 8	0.815 6
3	2.039 0	1.196 1
4	27.042 9	2.163 8
5	50.870 3	4.440 8

的优势;ANN模型的MSE值在5组中均为最低,第5组甚至达到了0.000 1。从这两个指标来看,ANN模型的拟合效果全面优于其他5种分布模型。从 $-2\ln L$ 这一指标来看,ANN模型在第3、5组数据集拟合效果最佳,Fréchet模型在第1、4组数据集上拟合效果最佳,Weibull模型在第1组数据集上拟合效果最佳(图3b)。

传统统计分布的故障率函数形式较为单一,由表5形状栏可知,Exponential分布故障率为常数;Weibull、Normal、Logistic分布故障率只表现为单调递增或递减;Fréchet分布的故障率形状更多表现为单峰曲线,这限制了统计分布函数的应用。而且真实世界环境复杂多变,传统模型无法描述U型、单峰型、双峰型甚至更复杂的故障率函数。观察图4的FRF栏可发现ANN模型更加灵活,当传统统计分布模型有着较好的拟合效果时,ANN模型不仅有着与传统模型整体一致的故障率函数趋势,还能体现更细微的故障率函数变化,在局部表现可表现为U型或单峰型故障率曲线,这也是ANN模型拟合精度更高的原因之一。

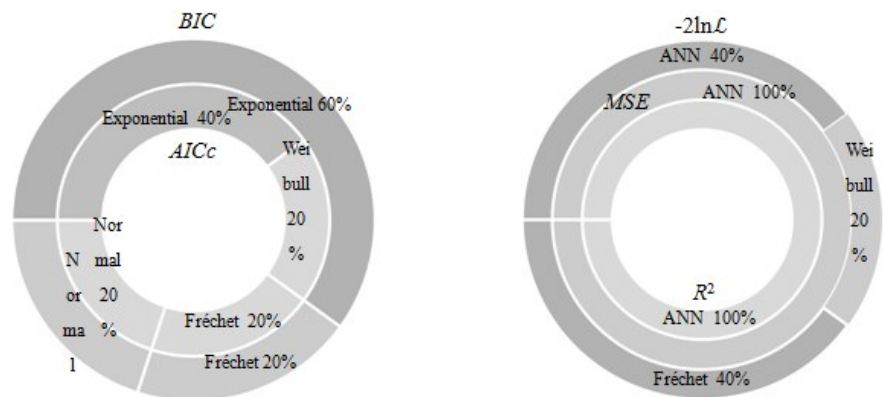
4 结 论

本文提出一种基于ANN的可靠度预测方法,以ReLU函数作为激活函数,以MSE作为回归损失函数,以故障时间和其中位秩估计值训练网络,在故障数据集时间范围内选取大量测试点并将其输入训练好的网络得到CDF预测值,然后利用样条回归法拟合出平滑完整的CDF曲线并推出PDF和CDF曲线。最后,结合婴儿培养箱等5组故障数据集进行验证,用 R^2 、MSE及 $-2\ln L$ 与Weibull、Exponential、Fréchet、Logistic和Normal分布进行对比,验证ANN模型的有效性和估计精度的优良性。综上所述,ANN模型能够从故障数据集中找出更准确的规律,能促进更有效的事事故预防。下一步将研究基于ANN模型的复杂系统剩余寿命预测。

表 5 ANN、Weibull、Exponential、Fréchet、Logistic 和 Normal 模型的拟合优度的参数值和故障率形状
Table 5 Failure rate shapes and goodness of fit of ANN, Weibull, Exponential, Fréchet, Logistic and Normal models

序号	数据集	模型	R^2	MSE	AIC_c	$-2\ln L$	BIC	形状
1	1	ANN	0.990 4	0.000 8	-	162.328 4	-	↓∩U
2	1	Weibull	0.979 0	0.001 7	166.070 1	161.14 70	166.692 2	↓
3	1	Exponential	0.867 7	0.010 5	165.570 3	163.284 6	166.057 2	-
4	1	Fréchet	0.984 2	0.001 2	165.757 3	160.834 2	166.379 4	∩
5	1	Logistic	0.866 4	0.010 6	187.894 2	182.971 1	188.516 3	↑
6	1	Normal	0.744 6	0.020 2	193.998 8	189.075 7	194.620 9	↑
7	2	ANN	0.970 4	0.002 4	-	154.594 7	-	∩U
8	2	Weibull	0.966 3	0.002 7	141.744 8	137.381 2	144.548 2	↓
9	2	Exponential	0.936 6	0.005 2	142.007 2	139.889 6	143.473 1	-
10	2	Fréchet	0.901 7	0.008 0	153.259 3	148.895 7	156.062 7	↓
11	2	Logistic	0.948 6	0.004 2	175.963 9	171.600 3	178.767 3	↑
12	2	Normal	0.905 5	0.007 7	179.613 0	175.249 3	182.416 4	↑
13	3	ANN	0.987 6	0.001 0	-	64.105 2	-	↑
14	3	Weibull	0.958 7	0.003 3	70.279 8	65.574 0	71.565 4	↑
15	3	Exponential	0.930 7	0.005 5	68.636 9	66.414 6	69.410 4	-
16	3	Fréchet	0.809 9	0.015 2	88.142 1	83.436 2	89.427 7	∩
17	3	Logistic	0.882 1	0.009 4	75.444 9	70.739 1	76.730 5	↑
18	3	Normal	0.978 1	0.001 7	74.914 7	70.208 8	76.200 3	↑
19	4	ANN	0.991 2	0.000 7	-	128.789 2	-	∩
20	4	Weibull	0.951 2	0.003 9	126.470 2	121.547 2	127.092 3	↑
21	4	Exponential	0.666 7	0.026 3	135.745 9	133.460 2	136.232 8	-
22	4	Fréchet	0.984 8	0.001 2	123.262 7	118.339 6	123.884 8	∩
23	4	Logistic	0.969 5	0.002 4	128.307 5	123.384 4	128.929 6	↑
24	4	Normal	0.919 0	0.006 4	129.262 0	124.338 9	129.884 1	↑
25	5	ANN	0.998 7	0.000 1	-	379.463 2	-	∩U
26	5	Weibull	0.991 8	0.000 7	389.536 6	385.281 3	393.105 3	↑
27	5	Exponential	0.333 5	0.054 6	485.921 0	483.837 7	487.749 7	-
28	5	Fréchet	0.974 9	0.002 1	394.688 2	390.432 9	398.257 0	∩
29	5	Logistic	0.994 6	0.000 4	389.715 9	385.460 6	393.284 6	↑
30	5	Normal	0.994 6	0.000 4	387.854 9	383.599 6	391.423 7	↑

↓表示单调递减;↑表示单调递增;-表示常数;U表示浴盆形;∩表示单峰型



a:根据BIC和 AIC_c 得到的最佳模型百分比环形图 b:根据 $-2\ln L$ 、MSE和 R^2 得到的最佳模型百分比环形图

图 3 根据拟合评估指标得出的最佳模型百分比

Figure 3 Percentages of the best models according to the fit quality criteria

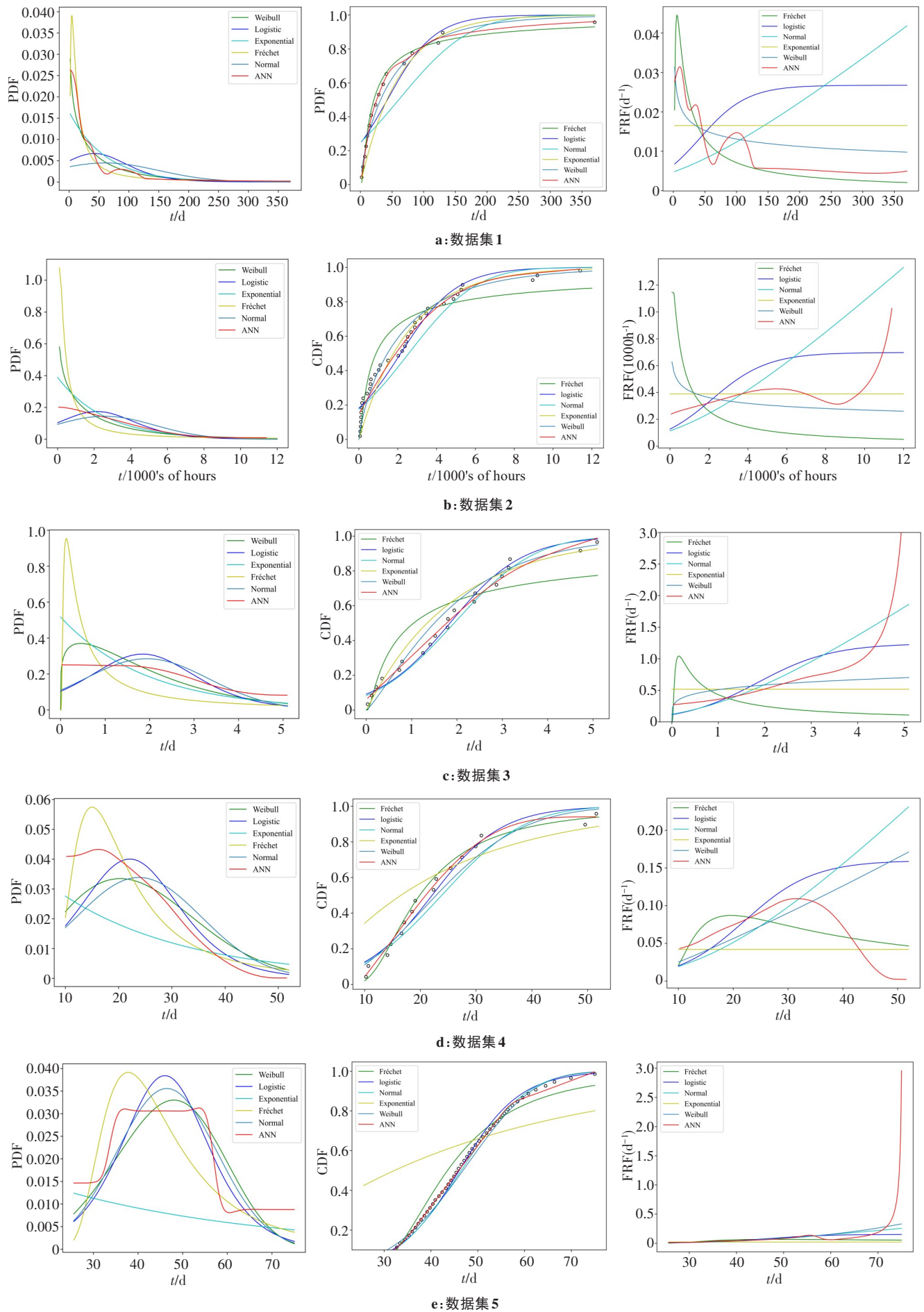


图4 ANN、Weibull、Exponential、Fréchet、Logistic和Normal模型的PDF、CDF和FRF

Figure 4 PDF, CDF and FRF of ANN, Weibull, Exponential, Fréchet, Logistic and Normal models

(下转 243 页)