

信息熵在缺陷漏磁信号量化中的应用

戴 光,赵海龙,杨志军

(东北石油大学 机械科学与工程学院,大庆 163318)

摘 要:缺陷漏磁信号的量化问题是漏磁信号研究领域中的一个重要方向。对缺陷漏磁信号进行FFT变换,根据变换后缺陷信号的幅值谱序列,对漏磁信号的信息熵进行了特征提取。利用提取的信号特征量,结合BP网络,完成了缺陷信号长度、深度的量化,量化精度分别达到了88.48%和84.17%。理论分析与试验研究证明了信息熵应用于漏磁信号量化过程中的可行性。

关键词:漏磁检测;信息熵;量化;BP网络

中图分类号: TG115.28

文献标志码: A

文章编号: 1000-6656(2011)02-0001-04

The Application of Information Entropy in Quantization of Magnetic Flux Leakage Signals of Defect

DAI Guang, ZHAO Hai-Long, YANG Zhi-Jun

(Mechanical Science and Engineering College, Northeast Petroleum University, Daqing 163318, China)

Abstract: The quantization of defect magnetic flux leakage signal is always an important direction in the field of magnetic flux leakage signal processing. In this paper, Fourier transform was applied to magnetic flux leakage signals. By the use of the amplitude spectral sequence of defect magnetic flux leakage signal, information entropy was extracted. The length and depth of defect were quantified through the combination of signal characteristic and BP network, and the quantization precision was estimated reach 88.48% and 84.17%. It was proved that information entropy was feasible in the quantization of defect magnetic flux leakage signal by theoretical analysis and experimental study.

Keywords: Magnetic flux leakage testing; Information entropy; Quantization; BP network

随着我国经济的发展和国家战略原油储备体系的建立,石油化工行业的储罐数量和容量迅猛增加,其安全性也得到人们的高度关注。漏磁检测以其具有易于实现自动化、可靠性较高等优点,在储罐底板检测中得到了广泛的应用。对检测做出评定的基础是漏磁信号,当今对漏磁信号的研究主要分为正问题与逆问题两个方向。漏磁检测的正问题是根据缺陷特征参数来评估漏磁场的大小;而逆问题则是指由漏磁信号包含的信息量化出缺陷的特征参数。

信息熵理论自从提出以来以其独特的优势在各个领域得到了广泛的应用。笔者以漏磁信号为基础,对检测得到的缺陷信号进行熵谱分析,并将缺陷信号的信息熵作为漏磁信号的一个特征量,用于漏磁信号的量化。最后通过理论分析与试验研究,证明了熵谱分析在漏磁信号量化过程中的可行性。

1 信息熵

信息是对事物状态的存在方式和相互联系进行描述的一组文字符号、语言、图像或情态。人们通过掌握的信息可以减少对客观事物的不确定程度。因此美国数学家、信息论的创始人 Shannon 曾在题为《通讯的数学理论》的论文中指出:“信息是用来消除随机不定性的东西”。可见信息与熵有着紧密的联系。1957 年 Jaynes 对信息熵做了如下定义^[1]:

收稿日期:2010-05-31

基金项目:全国高校博士点青年教师基金资助项目(20092322120001);黑龙江省自然科学基金资助项目(A200907);黑龙江省自然科学基金资助项目(E201016)

作者简介:戴光(1954—),男,博士,教授,主要从事现代无损检测和安全评价的研究。

假设某事件的可能结果和出现的几率如下:

$$\begin{array}{ll} X_1, X_2, \dots, X_n & \text{可能结果} \\ P_1, P_2, \dots, P_n & \text{出现的几率} \end{array}$$

且
$$\sum_{i=1}^n P_i = 1$$

则相应的信息熵为:

$$S = -K \sum_{i=1}^n P_i \ln P_i \quad (1)$$

式中 K 为一比例系数。

若每个结果出现的几率相同:

$$P_1 = P_2 = \dots = P_i = \dots = P$$

则式(1)变为:

$$S = -k \ln P \quad (2)$$

式(2)中 P 表示任意结果出现的几率,若用 Ω 表示可能出现的总数或微观状态数,且 $P=1/\Omega$,则式(2)转化成 $S=k \ln \Omega$,其中 k 称为玻尔兹曼常数,此时的熵往往被称为玻尔兹曼熵。当 K 取 $1/\ln 2$ 时, S 的单位为 bit,此时的信息熵为:

$$\begin{aligned} S &= -K \sum_{i=1}^n P_i \ln P_i = -\frac{1}{\ln 2} \sum_{i=1}^n P_i \frac{\log_2 P_i}{\log_2 e} = \\ &= -\sum_{i=1}^n P_i \log_2 P_i \end{aligned} \quad (3)$$

式(3)与 Shannon 提出的信息熵形式相同,故往往称此时的熵为 Shannon 熵。由上式可知,事情的可能性结果越多,每个结果出现的几率就越小,熵就越大,对事情的现状及发展变化越是难以判断。因此说,熵是无知度的度量。同时还可以发现,在 Ω 一定的情况下,当 $P_1 = P_2 = \dots = P_i = \dots = P$ 时,事物的不确定程度最大,此时的信息熵也最大,对应的熵值为:

$$S_{\max} = \log_2 \Omega \quad (4)$$

实际应用中,常常将式(3)与(4)之比作为信息的又一个度量,称之为相对信息熵,以 H 表示:

$$H = \frac{S}{S_{\max}} \quad (5)$$

2 漏磁信号的信息熵

在现实应用中,常常根据信息熵的来源不同,将相应的信息熵分为不同种类,如由小波分解得到的信息熵称为小波熵;由 FFT 频谱分析得到的信息熵称为谱熵。而根据 P 的含义不同,又可将信息熵进行进一步分类,如同样由 FFT 分析得到的信息熵,当 P 反映 FFT 功率谱分布时则称对应的谱熵为功率谱熵;当 P 反映信号经过 FFT 变化后的幅值谱

分布时,则称此时的谱熵为幅值谱熵。对应的相对熵则往往称为相对功率谱熵与相对幅值谱熵。2004 年王太勇将相对功率谱熵引入到漏磁信号识别过程中,通过试验证明该特征量可以消除速度对漏磁信号的影响,进而可以区分不同种类的缺陷^[2]。考虑到缺陷信号的相对幅值谱熵抗噪能力强,同时随缺陷尺寸的变化规律稳定且明显,笔者将缺陷信号的相对幅值谱熵^[3]作为漏磁信号的一个特征量,用于漏磁信号的量化处理。该特征量计算过程为:

(1) 确定缺陷信号的存在区间。

(2) 将分析信号的基线置零:可以通过 $s' = s - \bar{s}$ 获得,其中 s 为原始信号, \bar{s} 为 s 的平均值, s' 为置零后的信号。

(3) 对置零后的信号 s' 进行 FFT 变换(如图 1 和 2),设经过 FFT 变换后得到信号的频谱序列为 \hat{s} ,其中元素 \hat{s}_j 对应的频谱为 X_j 。

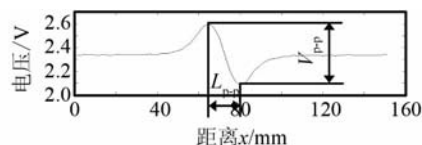


图1 原始信号及特征参量

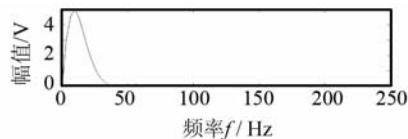


图2 原始信号的频谱

(4) 计算各个采集点对应的幅值谱熵:① 计算各频谱的模,记为 q_j ,则 $q_j = |X_j|$ 。② 计算各幅值的总和 $q_s = \sum_{j=1}^N q_j$ 。③ 令 $\hat{q}_j = \frac{q_j}{q_s}$,则该点对应的幅值谱熵为 $S_j = -\hat{q}_j \times \log_2(\hat{q}_j)$ (当 $\hat{q}_j = 0$ 时计 $S_j = 0$)。

(5) 求信号总的幅值谱熵:

$$S_s = \sum_{j=1}^N S_j$$

(6) 计算信号的相对幅值谱熵:

$$H = \frac{S_s}{\log_2 N}$$

式中 N 为分析信号的长度。

当被测底板有缺陷时,漏磁信号的谱图往往会在某些频率或频段处出现能量集中。相对幅值谱熵反映了漏磁信号幅值谱的集中程度,熵值越小,幅值越集中;熵值越大,幅值越分散。在漏磁检测中,腐蚀缺陷越小,对应的缺陷信号越难从原始信号中区

分出来,也就是说腐蚀缺陷尺寸越小,信号的不确定程度越大,信号越混乱,反映在熵领域中就是熵值越大;反之,腐蚀缺陷越大,缺陷信号越容易从原始信号中区分出来,不确定程度越小,对应的熵值也越小。所以可以用信号的相对幅值谱熵作为量化缺陷信号的一个特征量。

3 试验研究

定义缺陷的深度方向为垂直于平板的方向,缺陷的长度方向为检测仪的行走方向;量化的深度为缺陷的最大深度,量化的长度为沿检测仪行走方向各路传感器所对应的长度。为了便于信号的量化,笔者在该特征量的基础上又选择了缺陷信号的峰谷高值 V_{p-p} 和峰谷长值 L_{p-p} (图 1) 作为缺陷信号的另外两个特征量,分别用于量化缺陷的长度与深度。采用 BP 网络来完成缺陷的量化过程,量化过程的信息流量如图 3 所示。

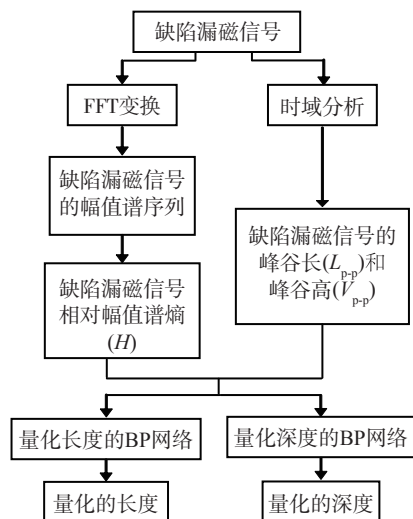


图 3 信号处理流程

网络的训练及仿真样本来源于试验标定板。标定板板厚为 6mm, 在标定板上人为加工了大小不等

的圆柱形缺陷 28 个, 其按直径分为 5, 10, 15, 20, 25, 30, 35 mm 共七组, 每组有 1, 2, 3, 4 mm 深四个圆柱形缺陷。利用 30 通道漏磁扫描检测仪对标定板进行检测。用 Matlab 软件按传感器通道对插值后的数据进行特征提取, 最后可得 595 个样本。将其中的 399 个样本用于网络的训练, 剩下的 196 个样本用于网络的仿真。网络训练的最大误差设定为 1×10^{-5} , 笔者采用两个双隐层的 BP 网络分别量化缺陷的长度与深度, 利用“试凑法”来确定中间隐层单元的神经元个数。最后确定深度方向的神经网络结构为 $3 \times 12 \times 15 \times 1$, 长度方向的神经网络结构为 $3 \times 10 \times 11 \times 1$ 。此时长度方向与深度方向神经网络最低仿真精度分别为 93.62% 与 97.60%。

对训练好的网络进行测试, 测试数据来源于试验板, 试验板的缺陷分布如图 4 所示。表 1 和 2 分别为两个网络的测试结果。

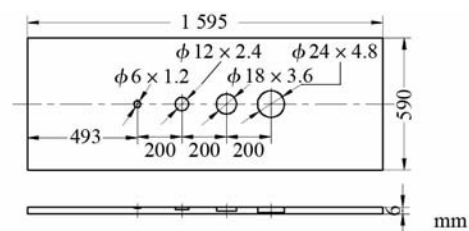


图 4 试验板缺陷分布

由表 1 和表 2 可知, 用该网络对缺陷进行量化, 深度方向的精度最低可达 84.17%, 长度方向的精度最低可达 88.48%。不难发现, 网络的测试精度明显低于相应的仿真精度。这是因为通过训练, 网络已经很好地掌握了标定板上缺陷与信号之间的对应关系, 但由于制造、加工、测量等误差的存在以及来自平板本身(如表面粗糙度)、外界环境(如电磁波干扰)、人为等因素的影响, 该对应关系在不同板上的表现形式存在差异。因此当用同一块板上的缺陷样本对网络进行仿真时, 网络的量化精度很高; 而用

表 1 缺陷深度量化结果

实际深度/mm	预测深度/mm	精度/%	实际深度/mm	预测深度/mm	精度/%	实际深度/mm	预测深度/mm	精度/%
1.20	1.19	99.17	2.40	2.62	90.83	3.60	3.81	94.17
1.20	1.20	100.00	3.60	3.54	98.33	3.60	4.10	86.11
2.40	2.51	95.42	3.60	3.03	84.17	3.60	3.80	94.44
2.40	2.48	96.67	3.60	4.09	86.39	4.80	4.44	92.50
2.40	2.71	87.08	3.60	3.30	91.67	4.80	4.70	97.92
2.40	2.31	96.25	3.60	3.73	96.39	4.80	4.76	99.17
2.40	2.58	92.50	3.60	3.81	94.17	4.80	4.81	99.79

表2 缺陷长度量化结果

实际长度/mm	预测长度/mm	精度/%	实际长度/mm	预测长度/mm	精度/%	实际长度/mm	预测长度/mm	精度/%
5.81	5.81	100.00	5.23	5.24	99.81	17.10	17.08	99.88
5.20	5.19	99.81	10.49	10.49	100.00	13.07	11.72	89.67
7.20	7.38	97.50	14.05	14.07	99.86	8.71	8.22	94.37
10.39	10.35	99.62	16.19	16.32	99.20	10.85	12.10	88.48
11.76	11.71	99.57	17.43	16.70	95.81	16.41	16.41	100.00
11.94	11.98	99.66	17.97	18.09	99.33	19.76	21.81	89.63
11.00	10.98	99.82	17.86	17.40	97.42	21.94	22.32	98.27

另一块板上的缺陷样本去测试网络性能时,网络的量化精度会下降,但该精度仍在可接受的范围之内。故可以将信息熵作为缺陷量化的一个特征量。

4 结论

(1) 对缺陷漏磁信号进行了 FFT 变换,并根据变换后的幅值谱序列完成了缺陷漏磁信号信息熵的提取。

(2) 建立了用于量化缺陷深度与长度两个方向的 BP 网络,并根据标定板数据完成了神经网络的训练与仿真。长度与深度两个方向的 BP 网络最低仿真精度分别达到了 93.62% 和 97.60%。

(3) 根据试验板数据对网络的性能进行了测

试,深度方向神经网络的预测精度最低可达 84.17%、长度方向神经网络的预测精度最低可达 88.48%,故可以将缺陷信号的信息熵作为漏磁信号量化的一个特征量,用于漏磁信号的量化处理。

参考文献:

- [1] 李鹤龄. 信息熵、玻尔兹曼熵以及克劳修斯熵之间的关系——兼论玻尔兹曼熵和克劳修斯熵是否等价[J]. 大学物理, 2004, 23(12): 37-40.
- [2] 王太勇, 刘兴荣, 秦旭达, 等. 谱熵分析方法在漏磁信号特征提取中的应用[J]. 天津大学学报, 2004, 37(3): 216-220.
- [3] 刘红星, 左洪福, 姜澄宇, 等. 信号频谱的二维向量及其应用[J]. 中国机械工程, 1999, 10(5): 537-539.

关于 2011 年度航空航天无损检测认证部(NANDTB-CN) 举办 NAS410 资格鉴定考试报名的通知

航空航天无损检测认证部(NANDTB-CN)是由国防科技工业无损检测人员资格鉴定委员会(DiNDT)成立,提供满足 NAS410/EN4179 国际宇航准则的无损检测人员资格符合性认证服务。目前 NANDTB-CN 已有 41 家成员单位 61 名代表,包括了国内主要航空航天领域的企事业单位(CAAC, COMAC, AVIC, CASC, CASIC 等)及来自美洲、欧洲及亚太地区企业(GEAE, Goodrich, Honeywell, Rolls-Royce, SAM 等)的代表,是 Nadcap 审核机构 PRI 在中国地区的唯一认可机构。

根据 NANDTB-CN 全体成员大会制定的 2011 年度工作计划的安排,拟定于 2011 年 6 月,在西安

举办符合 NAS410/EN4179 要求的无损检测五个专业(新增加涡流检测 ET 专业)的 II, III 级人员资格鉴定考试。请各有关单位按照要求报名。考试时间、地点等具体安排待报名资格审查通过后另行通知。

相关报考说明、资格鉴定申请表、航空航天典型代表试件分类表以及实际操作考试项目分类表请与鉴定委秘书处联系。联系方式如下:

联系人: 张京焘, 秦学

联系电话: 010-84380385, 84387131

传 真: 010-64615252

电子邮箱: nandtb@163.com

欢迎网上投稿 欢迎订阅 欢迎刊登广告