学

갚

志

Vol. 24 No. 1 Mar. 2007

文章编号: 1000-4556(2007)01-0043-09

波

谱

基于神经网络的 NMR 谱图自动相位 校正综合算法

黄颖颖,李 鹏,刘小征,杨 光*

(上海市功能磁共振成像重点实验室(华东师范大学),华东师范大学物理系,上海 200062)

摘 要:目前,已有多种算法被应用在核磁共振谱图自动相位校正中,由于各种算法本身特 性和所基于谱图的具体特性的差异,不同算法对于特征不同的谱图的适用性也各不相同.针 对这一情况,本文在综合研究多种现有自动相位校正算法的基础上,提出了一种基于神经网 络的,可以根据谱图的特征来选取最合适的算法进行自动相位校正的综合算法.实验表明, 本算法可以获得比以往方法更好的计算结果.

关键词:核磁共振;自动相位校正;神经网络;综合算法;NNAPC 中图分类号:O482.53 文献标识码:A

引言

傅立叶变换后得到的核磁共振(NMR)谱常常存在相位偏差,峰形表现为吸收型和 色散型的混合线型^[1].为了得到相位正确的谱图,就需要对谱图进行相位校正.相位校 正通常通过交互的方法手工进行,但是随着谱仪采样和数据处理自动化程度的提高,自 动相位校正的应用也越来越广.

相位校正的过程可以用以下公式来描述:

$$\begin{aligned} R_i &= R_i^0 \cos(\phi_i) - I_i^0 \sin(\phi_i) \\ I_i &= I_i^0 \cos(\phi_i) + R_i^0 \sin(\phi_i) \end{aligned} \tag{1}$$

其中 R_i^o 和 I_i^o 分别是未经相位校正的谱的第 *i* 个点的实部和虚部; R_i 和 I_i 是相位校正之 后的第 *i* 个点的实部和虚部; ϕ_i 是对第 *i* 个点进行相位校正时的相位角, 它可表示为:

收稿日期: 2006-05-23; 收修改稿日期: 2006-06-14

基金项目:国家自然科学基金资助项目(10575038)

作者简介:黄颖颖(1980-),女,浙江人,硕士,无线电物理专业. *通讯联系人:杨光,电话:021-62233293, E-mail: gyang@phy. ecnu. edu. cn.

(2)

 $\phi_i = phc0 + phc1 \times i/N$

其中 $phc0 \ge 0$ 级相位校正项; $phc1 \times i/N \ge 1$ 级校正项, $phc1 \ge R$ 数, N 是谱上总数 据点数. 所谓自动相位校正, 就是要根据谱图的特征, 来获得 phc0 和 phc1 这两个常数, 从而利用上述公式完成对谱图的相位校正.

自 1969 年 Ernst^[2]提出第一个自动相位校正算法到现在,已出现了多种自动相位校正算法.根据这些算法的特点,可以大致把它们分为两大类:

第一类算法总是利用整谱的特性,设计一个在吸收谱型中取得最大或最小值的目标 函数,并找出一组使得上述目标函数最优化的 phc0 和 phc1 数值.由于目标函数通常是 非线性的,这些算法通常都使用迭代算法.这类算法常用的判据有.利用吸收线型的实 部积分最大和虚部积分为零的特点^[2-4]、基线最优化^[5]、ZOE 零级平均^[6]、ACME 熵最 小化^[7].这类算法通常受基线、噪声的影响比较大.

第二类算法则首先利用单个吸收峰的特点,确定谱图中个别谱峰的相位偏角,然后 通过线性化的方法得到整张谱图的 phc0 和 phc1. 这类算法的关键步骤就是单个谱峰相 角的计算,常见的算法有:基于时域的算法、拟合成洛伦兹线形算法、基于 DISPA 圆 的算法^[10-13]、APSL 置对称性算法^[14,15]、PAMPAS 算法^[16].这些算法各有其不同的特 点,例如:Heuer 提出的 APSL 置对称性算法^[14],利用了吸收线型具有最大的对称性的 特点,该算法受基线畸变的影响小,但是对噪声、谱峰位置测量的准确性以及峰型的不 对称则比较敏感,同时要求谱图有足够高的数字分辨率.王超等人提出的基于 DISPA 圆 的 CLESP 等密度点连线法^[13],能够较好地处理基线畸变、谱峰重叠和选取的峰型不完 整等情况,但在信噪比很低的情况下计算结果误差较大.Dzakula Z 提出的 PAMPAS 算 法^[16],算法思想简洁,但要求取点必须包括完整的峰型并需要准确测定谱峰位置,必要 时需要进行局部基线校正.Sotack 等人提出的基于 DISPA 圆线形分析的算法^[14],算法 计算简单、速度快,但是对基线畸变、重叠峰比较敏感.

需要指出的是, 第一类算法中的判据大多可以用在第二类算法中计算单个谱峰的相 位角, 在这种情况下, 单个谱峰范围内的一级相位偏差一般可以忽略.

由于不同算法计算结果的准确性受不同因素(如信噪比、分辨率、对称性、谱峰重叠 程度、基线畸变的谱峰等等)的影响各不相同,因而,如果能够根据具体的谱峰的特征, 选择出最适合该谱峰的自动相位校正方法来进行相位校正,就能够获得最优的结果.本 文的工作就是利用人工神经网络来对谱峰的模式进行判别,根据具体的谱峰模式来选择 最适用于该谱峰的相位校正算法,以取得最优的计算结果.

1 理论

神经网络系统是一种高度自适应的非线性系统,它可以通过对样本的学习训练抽取 出隐藏在样本中的因果关系,在模式分类中有着广泛的应用.以本文中所用的三层误差 反向传输(back propagation, BP)人工神经网络为例,其结构通常包括输入层、输出层和 一定数量的中间层(见图 1).不同层的节点(或称神经元)之间,通过一定的权重相连接. 除了输入层外,每个节点的输入值,就是所有与之相连的上层节点的加权和.

如图1所示,以中间隐藏层为例,隐藏层节点的输出值可用下面公式表示:

$$Q_j = f(\sum X_i \times W_{ji} + b_j) \tag{3}$$

其中, X_i 是输入层节点值; W_{ji}是连接从输入到隐藏层的权值, O_j 是隐藏层节点输 出值. b_j 是节点阈值, 可以看作一个输入恒为1的网络权值. 网络权值 W_{ji}和阈值 b_j 都 是可调的, 正是基于神经网络权值和阈值的动态调整, 节点乃至神经网络才得以表现出 某种行为特性. *f* 是节点的传递函数, 用于对上述求和单元的计算结果进行函数运算, 得到隐藏层节点的输出, 常用的函数形式有线性和非线性 S 型.

所以整个运算过程可以简述为:输入信号经神经元加权求和及传递函数作用后,得 到最终的输出.



图 1 三层 BP 神经网络模型 Fig. 1 A typical BP neural network structure

神经网络的应用过程,先是利用一批已知输入输出的样本对神经网络进行训练,即 使用迭代的非线性优化算法,调整神经网络中连接不同层节点的权重,使得特定输入下 的输出和已知的输出尽可能一致.这一步骤速度通常较慢.在训练完成后,就可以利用 训练好的神经网络对未知样本进行预测,即利用已知的神经网络的权重,根据输入和公 式(3)计算输出结果,这一步骤只涉及简单的乘加运算,因此速度很快.

理论已证明, 三层 BP 神经网络, 只要隐藏层节点数足够多, 就具有模拟任意复杂 的非线性映射的能力. 但在应用神经网络解决实际问题的过程中, 关于网络的层数、每 层的神经元节点数、传输函数、训练算法的选择, 并无成熟的理论指导, 主要是通过大 量的实验尝试获得.

本文中提出的算法中利用神经网络的目的,就是希望神经网络能够利用谱图的一些 特性作为输入,自动选取最适合特定谱峰的相位角计算方法,因此我们将算法命名为基 于神经网络的自动相位校正综合算法(Neural Network based Automatic Phase Correction, NNAPC). 在我们的算法中采用了三层 BP 神经网络,经过实验,中间层神经元的 数目选择为 7. 以下就算法中涉及到的神经网络的输入、输出等问题分别进行论述: NNAPC 算法是期望利用神经网络来根据谱峰的特点选择最适用的计算谱峰相位偏角的算法,因此我们选择实现了第二类算法中的4种代表性方法,即PAMPAS 算法^[16]、 TRIANGLE 算法^[11]、APSL 算法^[14,15]、CLESP 算法^[13].神经网络的输出应该直接标识 出某个选中的方法.本文采用多个输出节点的方式,如期望的输出值[1000]代表选中 了第一种方法,输出值[0100] 代表选中了第二种方法,依次类推.实际计算过程中, 神经网络的实际输出值不可能都是整数,所以我们将选择4个输出节点中输出值最大的 那个节点所代表的方法,例如,如果输出[0.80.10.30.1],我们就选择第一种方法.

和神经网络输出密切相关的一个概念是有关神经网络性能函数的选择,性能函数用 来在训练过程中描述计算输出和预期输出之间的误差,训练的过程实际上也就是性能函 数最小化的过程.较为常用的神经网络性能函数是均方差,以均方差作为性能函数的最 大问题在于它不能准确地反映不同的相位角计算算法之间的相对误差.由于神经网络的 训练并不能保证对于所有的训练样本都能够使实际输出符合期望输出,在这种情况下, 应该尽可能地使神经网络避免选择那些误差很大的算法.为此,我们定义性能函数为;

$$P(o) = \sum_{n=1}^{N} \sum_{i=1}^{4} | o_{ni} - o_{eni} | E_{ni}$$
(4)

其中 n 是对于所有的训练样本加和, i 是对于输出矩阵的 4 个节点求和, O_{ni}是第 n 个样本第 i 个节点的计算输出, O_{mi}是第 n 个样本的第 i 个节点的期望输出. E_{ni}是第 n 个样本的第 i 个节点所代表的相角计算算法的计算误差. 使用这一性能函数的直观结果 是使得神经网络在选择算法的时候,将尽可能地避开误差最大的那些算法.

1.2 神经网络的输入

在 NNAPC 算法中,神经网络的输入就是一些参数,用来表示影响算法表现的谱图的特征,如数字分辨率、信噪比等等.有很多因素都可能对不同算法的表现产生影响, 但能够作为神经网络输入的因素必须是那些可以从原始谱图中直接获得的因素,这样的 因素我们称为可评价因素.NNAPC 算法中使用的可评价因素包括数字分辨率、取峰范 围、信噪比和谱峰重叠程度.

数字分辨率越高,可供相位校正算法利用的点就越多,采样点就能越精确地描述真 实的谱峰,提供更高精度的积分面积、峰中心位置等信息,算法的计算结果就越准确. 本文采用半高宽范围对应的点数来表示数字分辨率.

取峰范围反映取峰的完整性,取峰范围越大,峰型越完整,对于利用线型进行计算 的算法有利;但由于取峰范围增大会引入低信噪比的点,对于抗噪声能力差的算法却不 利;此外,谱峰重叠的影响,也使得实际中往往不能选取完整的峰型进行计算.本文采 用半高宽的倍数来表示取峰范围.

信噪比反映信号和噪声的幅度比值, 信噪比越高, 数据就越精确, 所有的算法都能 给出更好的计算结果, 不过不同的算法在抗噪声能力方面有着明显的差异, 因此信噪比 对于算法的选择有着很大影响. 本文通过利用模谱信号的强度来计算谱峰的信噪比.

谱峰重叠程度同样会对算法计算的准确性产生影响,因为谱峰的重叠,会直接造成峰型的改变,也会改变峰顶位置的数值.较低程度的谱峰重叠的影响可以近似看作是低阶的基线畸变,谱峰重叠程度越高,算法受影响的程度就越大.我们用要研究峰与相邻

峰的距离和相邻峰与要研究峰的高度比2个参数来表示谱峰的重叠程度.

1.3 神经网络的使用

人工神经网络的使用大致可以分为 3 个阶段, 即训练、验证和应用. 训练就是根据 已知预期输出的样本, 来调整神经网络的权重的过程, 在 NNAPC 算法中, 训练过程的 步骤包括:

1.随机地产生一批具有已知谱图特征和相位偏差的谱峰,使用4种相位校正算法中的每一种对产生的谱峰进行相位偏角计算,得到每种算法对每个谱峰进行相角计算时的误差;

 2. 用上述方法产生的数据作为样本进行训练,其输入就是表示谱图特征的参数,而 预期输出则是表示最小误差方法的矩阵,训练时所使用的性能函数如公式(4)所示.

验证过程是采用分布特性和训练样本类似的独立样本,来检验经过训练的神经网络 的可用性. 其基本步骤包括:

1. 同训练步骤 1;

 2. 使用根据谱图数据提取出来的谱图特征参数作为输入,用神经网络计算每个样本 的输出矩阵,利用输出矩阵选择的算法,对具体样本的相位偏角进行计算;

3. 利用 1, 2 中的结果, 对神经网络选择的结果进行统计, 同时将结果的优劣和传统的方法进行比较;

实际应用的过程中,NNAPC 算法的总体步骤和大多数第二类相位校正算法类似, 基本步骤为:

1. 从谱图中选取适合进行相位偏角计算的谱峰,同时计算这些谱峰的谱峰特征参数;

2. 利用神经网络选择最适合每个谱峰的相位角计算算法,计算每个谱峰的相角;

3. 利用已知的谱峰的位置和相角进行线性化,计算出谱图的零级和一级相位校正系数,然后利用这两个系数和公式(1)~(3)对谱图进行相位校正.

2 实验与结果讨论

2.1 神经网络的训练

为了使人工神经网络能够处理比较普遍的情况,我们在产生训练样本的时候采取了 对谱图因素进行随机化的处理方法,对于产生的每个谱图样本,我们让每个谱图特征参 数都在一定范围内独立地随机变化,产生出来的谱峰再施加随机的相角偏差.按照理论 部分叙述的步骤,共产生 5 000 个样本作为神经网络的训练样本.各谱峰特征参数的取 值范围见表 1;

神经网络的结构,特别是中间层神经元的个数的确定,以及训练过程的参数的确 定,都是需要不断尝试的过程.我们在实验中发现,隐藏层节点数为7个时效果较好, 同时训练的时间也不至于过长,所以以下实验采用网络基本结构为5-7-4的BP网络.输 入层至隐藏层至输出层的传递函数分别取非线性S型函数(Sigmoid 函数)中的 tansig, logsig 函数,以完成输入模式到输出模式的非线性映射.学习算法采用 trainscg 学习算 法,固定训练迭代次数为1000次.

Table 1	Range of the characteristic parameters		
影响因素	取值范围		
数字分辨率	[2, 20]		
取峰范围	[2.0, 8.0]		
信噪比	[5.0, 100.0]		
与临峰间距	[3.0, 30]		
与临峰高度比	[0,0.8]		

表1 谱峰参数的取值范围

2.2 神经网络的验证

为了了解 NNAPC 算法的可用性,采用与产生训练样本同样的方法,产生了 10 万 个仿真测试样本用于算法测试.这些样本完全独立于训练样本,而且对于 NNAPC 算法 来说,它们的谱图参数和相位偏角都是未知的.NNAPC 算法首先根据谱图数据计算出 神经网络输入所需的谱图特征参数,然后利用神经网络选择的最合适的算法计算特定谱 峰的相位角.我们对 10 万个样本的计算结果进行了统计,结果如表 2 所示.

表 2 NNAPC 算法和所采用的 4 种算法计算结果比较

算法	算法为最优的 比率(%)	单纯使用某算法进行计算的 误差绝对值的均值(度)	单纯使用某算法进行计算的 误差绝对值的标准偏差(度)	
PAMPAS	2.396	5.40	4.27	
TRIANGLE	42.670	2.17	2.30	
APSL	29.015	2.31	2.24	
CLESP	26.522	7.91	19.55	
NNAPC	58.779	1.53	1.78	

Table 2 The results of NNAPC and four other algorithms

从表 2 中第二列可以看出,正如我们设想的,不同的谱图特征,确实会影响不同算 法的相对表现,在我们考察的 4 种算法中,在考察的影响因素变动范围内,没有哪一种 能够占据压倒性的优势.NNAPC 算法在 58.8%情况下挑选到了最佳的算法,说明了神 经网络运用于该模式识别问题的有效性.这个数字看起来似乎不高,但是和其他算法中 最有效的 Triangle 方法相比,仍然提高了 16.1%.需要指出的是,我们大量实验的结果 表明,如果在训练的时候单纯追求优选率的最大化,那么结果是相角计算误差的绝对值 的平均值会偏大.所以,我们通过选择目前所使用的性能函数的方法,来尽量避开误差 最大的算法,而不是在尽可能多的情况下选择最优的算法.我们的实验结果表明, NNAPC 算法选择到最差结果的比率约为 1.433%,说明我们选择的性能函数达到了预 期的效果.不同算法计算相位角的误差绝对值的平均值和标准偏差的结果表明, NNAPC 算法在所有的算法中取得了最小的计算误差和最小的误差标准差,说明 NNAPC 算法计算的准确性和算法的稳定性都优于使用任何一种单一算法的结果.

从表 2 中还可以看出, CLESP 算法在大约 1/4 的样本中表现出了最优的性能, 但其

误差绝对值的均值和标准偏差远高于其他算法,说明在我们考虑的影响因素变动的范围 内,存在着该算法根本无法适应的情况,这种情况大大影响了该算法的实用性.但是, 在 NNAPC 算法中,仍然可以有效地利用这种在局部情况下表现优越的算法来提高算法 整体的表现.可以说,NNAPC 算法提供了一种自动相位校正的新思路,只要已有的相 位角计算算法在特定的谱图条件下存在一些优势就可以被本算法采用,因而它也为其他 已有的自动相位校正算法提供了发挥作用的新的可能性.

2.3 实际实验得到的谱图数据分析及校正效果对比

图 2 是布鲁克核磁谱仪上得到的一张¹³C 去偶谱,未做相位校正.



图 2 核磁谱仪上得到的某样品¹³C去偶谱,未做相位校正

Fig. 2 Decoupled ¹³C spectrum acquired by a Bruker NMR spectrometer without phase correction

对上述¹³C去偶谱数据进行自动数据分析:先利用检峰程序从谱图中共选取出 10 个 适合进行相位偏角计算的谱峰,并利用谱特征提取方法自动提取出这些峰的谱峰特征参 数如表 3 中所示,然后将这些特征参数组成的特征向量输入前面已训练过的神经网络进 行计算,得到适合各个谱峰的相角计算算法.

表 3 实际谱图样本分析及经神经网络仿真结果

Table 3 Experimental spectrum analysis and the simulation result after using Nural Network

峰编号		\# 12 \# \+				
	数字分辨率	取峰范围	信噪比	与临峰间距	与临峰高度比	辸 拴昇 冮
1	25	3.88	135.10	2670.00	0.85	TRIANGLE
2	10	6.30	114.47	2547.00	0.25	TRIANGLE
3	9	4.33	28.98	14.67	3.04	APSL
4	8	8.88	88.05	16.50	0.33	TRIANGLE
5	7	15.29	183.56	19.00	0.48	APSL
6	7	18.43	213.63	19.14	0.84	CLESP
7	7	14.00	180.07	19.00	0.46	APSL
8	6	8.17	82.03	22.17	2.20	APSL
9	6	6.17	25.39	22.33	3.23	APSL
10	4	10.50	53.40	3818.00	0.48	APSL

图 3 显示了用 NNAPC 算法根据表 3 结果对图 2 所示¹³C 去偶谱进行自动相位校正的结果,可以看出本算法确实取得了令人满意的效果.需要指出的是,从表 3 可以看出,即使对于同一张谱图,由于每个谱峰的情况不尽相同,也应该选择不同的算法来进行相位角的计算.这也是 NNAPC 方法的一个优势.



图 3 用 NNAPC 算法对图 2 所示的¹³C 去偶谱进行自动相位校正结果图

Fig. 3 Decoupled ¹³C spectrum after NNAPC phase correction

3 总结

本文提出的 NNAPC 综合算法提出了自动相位校正算法的新思路,相比以前各种单一的算法,这种综合的算法可以充分发挥各种算法的长处,得到显著优于以往任何一种 算法的结果.即使对于同一谱图中的不同谱峰,也可以充分利用 NNAPC 的这一特点来 获得最优的结果.NNAPC 算法,为那些在某个方面表现极佳但整体表现可能欠佳的算 法提供了新的发展空间.

致谢: 感谢 Patrice Koehl 教授^[15] 在我们实现 APSL 算法过程中提供的帮助和关于 APSL 算法的改进建议.

参考文献:

- Qiu Zu-wen(表祖文), Pei Feng-kui(裴奉奎). Spectroscopy of Nuclear Magnetic Resonance(核磁共振波谱)
 [M]. Beijing(北京): Science Press(科学出版社), 1989. 202-203.
- [2] Ernst R R. Numerical hilbert transform and automatic phase correction in magenetic resonance spectroscopy[J].
 J Magn Reson, 1969, 1: 7-26.
- [3] Marshall M S. The use of modified simplex method for automatic phase correction in fourier-transform nuclear magnetic resonance spectroscopy [J]. Anal Chim Acta, 1981, 133: 103-108.
- [4] Chen C N, Kan L S. An iterative phase correction program for nuclear magnetic resonace(NMR) spectra[J]. Comput Meth Prog Bio, 1988, 26: 81-84.
- [5] Brown D E, Campbell T W, Moore Richard N. Automated phase correction of FT NMR spectra by baseline Optimization[J]. J Magn Reson, 1989, 85: 15-23.
- [6] Balacco G. A new criterion for automatic phase correction of high-resolution NMR spectra which does not require isolated or symmetrical lines[J]. J Magn Reson, Series A, 1994, 110: 19-25.
- [7] Chen Li, Weng Z Q, Goh L Y, et al. An efficient algorithm for automatic phase correction of NMR spectra based on entropy minimization[J]. J Magn Reson, 2002, 158: 164-168.

- [8] Backkuijsen H, Beer R D, Bovee W M J, *et al*. Retrieval of frequentcies, amplitude, damping factors, and phases from time-domain signals using a linear least-squares procedure[J]. J Magn Reson, 1985, 61: 465-481.
- [9] Montigny F, Elbayed K, Brondeau J, et al. Automatic phase correction of fourier transform nuclear magnetic resonance spectroscopy data and estimation of peak area by fitting to a lorentzian shape [J]. Anal Chem, 1990, 62: 864-867.
- [10] Craig E C, Marshall A G. Automated phase correction of FT NMR spectra by means of phase measurement based on dispersion versus absorption relation(DISPA)[J]. J Magn Reson, 1988, 76: 458-4751.
- [11] Sotak C H, Dumoulin C L, Newsham M D. Automatic phase correction of fourier transform NMR spectra based on the dispersion versus absorption(DISPA) lineshape analysis[J]. J Magn Reson, 1984, 57: 453-462.
- [12] Wachter E A, Sidky E Y, Farrar T C. Calculation of phase-correction constants using the DISPA phase-angle estimation technique[J]. J Magn Reson, 1989, 82: 352-359.
- [13] Wang Chao(王超), Huang Ying-ying(黄颖颖), Yang Guang(杨光). Automatic phase correction for NMR spectra-connecting line of equally spaced points (CLESP)(核磁共振谱图自动相位校正新方法-等密度点连线法)[J]. Chinese J Magn Reson(波谱学杂志), 2004, 21(4): 445-457.
- [14] Heuer. A new algorithm for automatic phase correction by symmetrizing lines[J]. J Magn Reson, 1991, 91: 241.
- [15] Koehl P, Ling C, Lefevre J F. Automatic phase correction of NMR spectra: statistic and limits[J]. J Chim Phys, 1995, 92: 1 929-1 938.
- [16] Dzakula Z. Phase angle measurement from peak areas(PAMPAS)[J]. J Magn Reson, 2000, 146(1): 20-32.

A Synthetical Algorithm for Automatic Phase Correction of NMR Spectra Based on Neural Network (NNAPC)

HUANG Ying-ying, LI Peng, LIU Xiao-zheng, YANG Guang*

(Shanghai Key Laboratory of Functional Magnetic Resonance Imaging (East China Normal University), Department of Physics, East China Normal University, Shanghai, 200062, China)

Abstract: A synthetical algorithm for automatic phase correction of NMR spectra based on neural network (NNAPC) is proposed. Taking into account the characteristics of the NMR spectrum or the resonance peak, the proposed algorithm uses an artificial neural network to choose the most appropriate algorithm to calculate the phase angles of a given peak. Statistical analysis demonstrated that the NNAPC algorithm is more accurate and stable than the existing phase correction algorithms referenced.

Key words: NMR, automatic phase correction, neural network, NNAPC

^{*} Corresponding author: Yang Guang, Tel:021-62233293, E-mail:gyang@phy.ecnu.edu.cn.