

PLS-ANN 算法-NIR 光谱非破坏性 Norvasc 药物有效成分的定量分析

吕慧英¹, 任玉林², 刘名扬^{3,4}

(1. 吉林大学计算机科学与技术学院, 2. 化学学院, 长春 130012;
3. 清华大学化学系, 北京 100084; 4. 辽宁出入境检验检疫局技术中心, 大连 116001)

摘要 采用偏最小二乘(PLS)结合人工神经网络(ANN)算法解析 Norvasc(络活喜)药片的近红外(NIR)漫反射光谱, 实现了对其中有效成分苯磺酸氨氯地平的非破坏定量测定. 设计了最佳的 PLS-ANN 模型, 分别讨论了最佳波长范围、导数光谱及输入层和隐含层节点数对预测结果的影响. 以 HPLC 法的测定结果作标准, 苯磺酸氨氯地平浓度预测值的相对误差 RE < 3.5%, 该方法可用于 Norvasc 药品实际生产中的质量控制.

关键词 偏最小二乘; 人工神经网络; 近红外漫反射光谱; 非破坏定量分析; Norvasc 药片

中图分类号 O657 **文献标识码** A **文章编号** 0251-0790(2007)05-0847-04

随着近红外(Near infrared, NIR)光谱及计算机软件技术的发展, NIR 光谱在制药工业中的定性、定量测定中显示了巨大的潜力, 包括从活性组分测定到固体药剂的非破坏性表征, 从生产过程各阶段的在线监控到原料和产品的快速鉴定^[1]. NIR 光谱法用于测定药物中有效组分浓度的文献较多^[2~5]. 由于 NIR 光谱谱带宽, 光谱重叠严重, 所以无法直接分辨, 需要用化学计量学方法进行解析^[1].

Norvasc(络活喜)是处方量最大的治疗高血压的药物, 其有效成分是苯磺酸氨氯地平, 每个片剂中的含量在 4.869~5.142 mg 之间, 测定精度要求高, 通常采用高效液相色谱法(HPLC)检测. 但该法仪器成本高, 操作过程繁琐, 分析时间长, 需要有机溶剂和试剂溶样. 采用 NIR 光谱分析, 可实现直接快速、非破坏性的定量分析, 且不需要任何有机溶剂和试剂^[1]. 采用通常的化学计量学方法分辨如此精度药品 NIR 光谱的误差较大. 人工神经网络(ANN)算法是一种重要的非线性拟合工具, 可以很好地解析复杂的 NIR 光谱^[6~8], 偏最小二乘(Partial least squares, PLS)法是使用最广泛的算法, 可以对 NIR 光谱进行全光谱解析^[1]. 本文采用 PLS 法, 选择 NIR 光谱的最佳波长, 用 ANN 算法解析 NIR 光谱, 采用 PLS-ANN 混合算法^[9,10]预测药品中苯磺酸氨氯地平的含量. 讨论了输入层节点数和隐含层节点数对含量预测结果的影响, 比较了用 NIR 光谱数据及一阶导数和二阶导数 NIR 光谱数据分别建立的 PLS-ANN 模型对预测结果的影响, 并以 HPLC 法测定值作为标准, 研究了本文方法的准确度.

1 实验部分

1.1 试剂与仪器

固体硫酸钡(分析纯); 总共 80 个批号 Norvasc 药品由大连辉瑞制药有限公司提供, 其中有效成分苯磺酸氨氯地平的含量采用 HPLC 法测定.

日本岛津 UV-3100 型紫外-可见-近红外分光光度计, 附有 ISR-3100 积分球.

1.2 实验方法

将 Norvasc 药片放在 ISR-3100 积分球样品槽内, 用硫酸钡粉末固定, 压平硫酸钡, 使其表面光滑; 选择光谱狭缝宽 12 nm, NIR 光谱的扫描波长范围为 1100~2500 nm, 每 1 nm 记录 1 个吸光度值, 每个样品扫描 2 次, 取其平均值, 即得到 80 个 Norvasc 药品的 NIR 光谱.

收稿日期: 2006-05-29.

联系人简介: 刘名扬(1968年生), 男, 博士, 高级工程师, 从事化学计量学和光谱分析化学研究.

E-mail: liumy@mail.tsinghua.edu.cn

1.3 数据处理及软件

每个样品的 NIR 光谱共有 1401 个吸光度值, 将 NIR 光谱吸光度数据用 PLS-ANN 模型进行处理. PLS 软件为美国 Thero Nicolet 公司 TQ Analyst 6.1.1 中的程序, ANN 软件采用 Matlab 6.5 (Mathworks, Inc., USA) 中的程序, 其它软件自行编制.

2 结果与讨论

2.1 HPLC 法分析 Norvasc 药片中的苯磺酸氨氯地平的浓度

采用 HPLC 法分析了 80 个批号 Norvasc 药片中的苯磺酸氨氯地平的含量 (4.869 ~ 5.142 mg), 以此作为 NIR 光谱法测定 Norvasc 药片浓度时的参考值, 也是期望得到的最佳预测值. 将 80 个批号药片随机分为 3 组: 训练集 (Train set) 含 62 个样品, 监控集 (Monitoring set) 含 8 个样品, 其余 10 个样品作试验集 (Test set).

2.2 Norvasc 药片的 NIR 光谱

实验得到 80 个批号 Norvasc 药片的 NIR 光谱图, 图 1 为随机选取的其中 3 个 NIR 光谱图. 可见, Norvasc 药片的 NIR 光谱严重重叠, 无法直接分辨, 必须应用化学计量学方法进行光谱解析, 才能实现分辨和浓度预测.

2.3 PLS 模型选择最佳波长范围

在本实验中, 每个样品的 NIR 光谱有 1401 个吸光度值数据. 用 ANN 法对 NIR 光谱进行解析时的计算量大, 输入的神经元 (输入节点) 不能太多, 否则网络运行的时间太长, 会造成运算无法收敛, 通常情况下, ANN 模型要求的输入节点为几个到几十个, 不易进行全谱处理. 对 NIR 光谱的 1401 个吸光度值进行数据筛选时, 通常采用分段或每间隔一定数据点取一个吸光度值, 这样容易造成有效信息的丢失. 我们利用 PLS 法可以对 NIR 光谱进行全光谱解析的特点, 选择最佳的 NIR 光谱波长范围, 完成数据压缩, 减少计算量.

Thero Nicolet 公司 TQ Analyst 6.1.1 中的 PLS 软件可以实现波长选择, 通过实验, 最后选择 NIR 光谱最佳波长范围为 2248 ~ 2300 nm, 此范围内共有 53 个波长点的吸光度数据, 即输入节点数是 53.

2.4 ANN 模型的选择和评价标准

ANN 模型选用三层向前网络, 即 1 个输入层、1 个隐含层和 1 个输出层. 网络输入层代表被预测样本的特征参数, 是以训练集样本在不同波长点处的吸光度值作为输入层, 通过在输入层设定的节点数进行输入 (不同的节点数对应着不同的波长点的吸光度值). 网络输出层输出 1 个预测的浓度结果, 由于 Norvasc 药片中只有苯磺酸氨氯地平 1 个有效成分, 所以 ANN 网络输出层选择 1 个节点 (即只有 1 个神经元).

在 ANN 模型的网络自学习过程中, 首先输入训练集 62 个样品的吸光度值数据及期望的含量值 (参考值), 然后设定参数: Net train param, Epoch = 0 ~ 2000, Goal = 0, Show = 50, Time = inf, Tr = 0.5. 根据传递函数并按照一定的学习规则自动调节网络各层之间的连接权重进行迭代, 以使网络的实际输出值与期望输出值相比较达到一定的精度要求 (即小于设定的最大允许误差值).

ANN 模型的预测能力采用监控集样本的相对标准误差 [RSE (%)] 来检验:

$$\text{RSE}(\%) = \left[\frac{\sum_{i=1}^n (c_{ip} - c_i)^2}{\sum_{i=1}^n c_i^2} \right]^{1/2} \times 100$$

式中, n 为监控集的样本数, 本文为 8, c_{ip} 为监控集中第 i 个样本中有效成分的预测含量值, c_i 为第 i 个样本中有效成分的参考含量值.

2.5 NIR 光谱及其一阶和二阶导数谱图的 ANN 模型的参数选择

分别对 80 个 Norvasc 药品的 NIR 光谱数据进行一阶和二阶导数处理, 在最佳波长范围 2248 ~

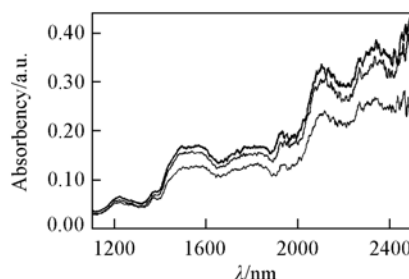


Fig. 1 NIR spectra of 3 batches Norvasc in tablets

2300 nm 内, 用 NIR 光谱的一阶、二阶导数数据(53 个波长点的数据)分别建立 ANN 模型, 并与原始 NIR 光谱建立的 ANN 模型进行对比. 分别预测监控集中 8 个样品的浓度值, 与参考值比较, 计算了 3 种光谱的相对标准误差 RSE(%), 有关参数见表 1.

Table 1 Selection of the vary ANN model parameters

| No. | Referenced concent by HPLC/mg | Conventional spectra | | First-derivative spectra | | Second-derivative spectra | |
|--------|----------------------------------|----------------------|--------|--------------------------|--------|---------------------------|---------|
| | | Calculated value/mg | RE(%) | Calculated value/mg | RE(%) | Calculated value/mg | RE(%) |
| 1 | 5.077 | 4.983 | -1.851 | 5.506 | 10.488 | 4.938 | -10.319 |
| 4 | 5.080 | 4.981 | -1.949 | 4.818 | -3.276 | 4.476 | -7.099 |
| 15 | 5.055 | 5.031 | -0.467 | 5.323 | 5.796 | 4.990 | -6.254 |
| 18 | 5.010 | 5.172 | 3.228 | 4.919 | -4.894 | 5.116 | 4.015 |
| 40 | 4.982 | 4.962 | -0.397 | 5.216 | 5.107 | 5.036 | -3.453 |
| 47 | 5.003 | 4.983 | -0.396 | 5.085 | 2.043 | 4.642 | -8.710 |
| 61 | 4.997 | 5.076 | 1.585 | 5.442 | 7.204 | 5.134 | -5.667 |
| 78 | 4.970 | 4.947 | -0.467 | 5.273 | 6.594 | 5.031 | -4.589 |
| RSE(%) | | 0.057 | | 0.191 | | 0.263 | |

由表 1 可见, 用原始 NIR 光谱数据建立的 ANN 模型的 RSE(%) 最小, 预测能力最好, 因此选择用原始 NIR 光谱数据建立 ANN 模型.

2.6 输入层节点数的选择

每个样品的 NIR 光谱在最佳波长范围内有 53 个吸光度值数据, 通过在 ANN 网络模型的输入层设定的节点数进行输入, 可以选择不同的波长间隔, 以判断可否再次筛选数据. 每间隔 2, 3, 5 和 7 nm 取一个吸光度值, 即网络输入节点数分别为 26, 17, 11 和 8 时, 分别预测监控集中 8 个样品的浓度, 与参考值比较, 统计其 RSE. 由图 2 可见, 不同输入波长间隔数(输入节点数)对测定结果的影响较大. 当输入节点数为 53 时, 监测集样品的相对标准误差 RSE% 最小, 所以选择 53 作为 ANN 模型的最佳输入节点数, 即数据压缩的过多易造成有效信息的丢失, 经过 PLS 法数据压缩后, 53 个波长点的吸光度值不宜再次压缩.

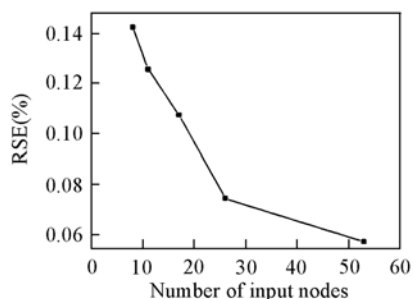


Fig. 2 Effect of the number of input nodes on RSE

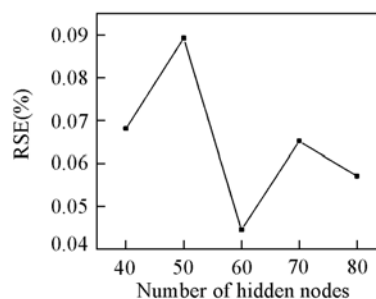


Fig. 3 Effect of the number of hidden nodes on RSE

2.7 隐含层节点数的选择

ANN 网络模型的隐含层神经元数目(隐含层节点数)对预测结果有较大影响. 在最佳输入波长范围内, 选择隐含层节点数分别为 40, 50, 60, 70 和 80, 其它参数均保持不变时, 预测监控集中 8 个样品的浓度, 分别统计其 RSE. 图 3 示出了不同隐含层节点数对预测结果的影响, 可见, RSE 曲线在节点数 50 处有最小值, 所以选择 50 作为最佳 ANN 模型的隐含层节点数.

2.8 试验集 Norvasc 药品分析

采用各优化参数建立了最佳的 ANN 模型. 用该 ANN 模型对试验集的 10 个 Norvasc 药品的原始 NIR 光谱进行解析, 分别预测其中的有效成分的浓度, 得到相应的苯磺酸氨氯地平预测含量值, 与 HPLC 法的测定值比较, 计算相对误差 RE(%). 为比较 PLS-ANN 混合算法的预测结果, 用单纯的 ANN 模型进行计算(在 1100 ~ 2500 nm 范围内取 27 个输入节点数, 隐含节点数选择为 50), 统计 RSE 结果见表 2. 由表 2 可见, PLS-ANN 混合模型的 RSE 比 ANN 模型小, 预测结果更准确; PLS-ANN 混合模型的 RE < 3.5%. 实验结果表明, PLS-ANN 模型对 Norvasc 药品 NIR 漫反射光谱具有较好的解析能力.

Table 2 Analytical results of the samples in the prediction set

| No. | HPLC Referenced | ANN | | PLS-ANNR | |
|--------|-----------------|---------------------|--------|---------------------|--------|
| | value/mg | Calculated value/mg | RE(%) | Calculated value/mg | RE(%) |
| 8 | 5.020 | 4.977 | -0.865 | 5.061 | 0.813 |
| 10 | 5.105 | 5.315 | 4.114 | 4.950 | -3.044 |
| 23 | 5.071 | 5.018 | -1.043 | 5.158 | 1.724 |
| 27 | 5.112 | 5.005 | -2.093 | 5.083 | -0.575 |
| 34 | 4.977 | 4.975 | -0.034 | 5.148 | 3.438 |
| 62 | 5.103 | 5.050 | -1.039 | 5.105 | 0.039 |
| 63 | 4.990 | 4.913 | -1.553 | 5.110 | 2.399 |
| 70 | 5.045 | 5.011 | -0.674 | 5.007 | -0.757 |
| 73 | 4.960 | 5.025 | 1.317 | 4.970 | 0.206 |
| 77 | 4.920 | 4.912 | -0.165 | 4.997 | 1.563 |
| RSE(%) | | 0.0886 | | 0.0688 | |

参 考 文 献

- [1] LU Wan-Zhen(陆婉珍), YUAN Hong-Fu(袁洪福), XU Guang-Tong(徐广通), *et al.*. Modern Analysis Technique of NIR(现代近红外光谱分析技术)[M], Beijing: Sinica Petro-Chemistry Press, 2000: 193—249
- [2] Morisseau K. M., Rhodes C. T.. Drug Development and Industrial Pharmacy[J], 1995, **21**(9): 1071—1078
- [3] Galante L. J., Brinkley M. A.. Anal. Chem. [J], 1990, **62**(23): 2514—2520
- [4] Ciurczak E. W.. Applied Spectroscopy Reviews[J], 1987, **23**(1/2): 147—154
- [5] LIU Ming-Yang(刘名扬), MENG Yu(孟昱). Chinese Journal of Spectroscopy Laboratory(光谱实验室)[J], 2006, **23**(6): 583—587
- [6] LIU Pei-Yi(刘培义), REN Yu-Lin(任玉林), GOU Yu-Hui(苟玉慧), *et al.*. Chem. J. Chinese Universities(高等学校化学学报)[J], 2000, **21**(4): 544—546
- [7] Ren Yu-Lin, Gou Yu-Hui, Ren Rui-Xue, *et al.*. Spectrosc. Lett. [J], 1999, **32**: 431—442
- [8] Liu Ming-Yang, Meng Yu, Li Jun-Feng, *et al.*. Chinese Chemical Letters[J], 2006, **17**(9): 1209—1212
- [9] CHEN Da(陈达), WANG Fang(王芳), SHAO Xue-Guang(邵学广), *et al.*. Spectroscopy and Spectral Analysis(光谱学与光谱分析)[J], 2004, **24**(6): 672—674
- [10] WANG Guo-Qing(王国庆), SHAO Xue-Guang(邵学广). Chin. J. Anal. Chem. (分析化学)[J], 2005, **33**(2): 191—194

Nondestructive Quantitative Analysis of Norvasc in Medicine Tablets by PLS-ANN Arithmetic Combined with NIR Spectroscopy

LÜ Hui-Ying¹, REN Yu-Lin², LIU Ming-Yang^{3,4*}

(1. College of Computer Science and Technology, 2. College of Chemistry, Jilin University, Changchun 130012, China;

3. Department of Chemistry, Tsinghua University, Beijing 100084, China;

4. Centre of Technique, Liaoning Entry-Exit Inspection and Quarantine Bureau, Dalian 116600, China)

Abstract Artificial neural network(ANN) algorithm combined with partial least squares(PLS) was used to parse near infrared(NIR) reflectance spectra of Norvasc medicine in tablets. The nondestructive and quantitative determination of the contents of amlodipine besylate, which is an effective component in the Norvasc tablets, were accurately carried out. The best model of PLS-ANN was designed. The effect of the best wavelength range, derivative NIR spectrum, input nodes and hidden nodes on the predicted results was discussed respectively. Compared the results with those of HPLC, the relative errors(RE) of mlodipine besylate are less than 3.5%. The analytical results could be applied to the quality control of Norvasc medicines in practical manufacture.

Keywords Partial least squares; Artificial neural network; Near infrared reflectance spectrum; Nondestructive and quantitative analysis; Norvasc in tablets (Ed.: K, G)