

中低分辨率光谱的恒星 Mg 元素丰度估计方法研究

卜育德^{1*}, 潘景昌², 衣振萍²

1. 山东大学(威海)数学与统计学院, 山东 威海 264209

2. 山东大学(威海)机电与信息工程学院, 山东 威海 264209

摘要 主要研究了一种新的基于 ELM 算法的中低分辨率光谱的恒星 Mg 元素丰度估计方法。大科学工程郭守敬望远镜(LAMOST)为我们提供了海量的中低分辨率的光谱, 确定这些光谱的 Mg 元素丰度将有助于我们深入了解银河系的形成历史和演化过程。目前从中低分辨率光谱中确定 Mg 元素丰度的方法主要是模板匹配法, 但该方法算法复杂, 优化参数较为困难且对噪声敏感, 因此有必要研究新的方法。实验结果显示, ELM 算法对 MILES 光谱的 Mg 丰度的估计的精度为 0.009 9(0.15)dex, 而对信噪比大于 50 的 LAMOST 光谱的精度为 0.002 7(0.11)dex。通过与其他算法进行对比, 证实 ELM 算法是一种能精确估计中低分辨率光谱的 Mg 元素丰度的算法, 能够应用于 LAMOST 后期的光谱数据中。

关键词 ELM 算法; Mg 元素丰度; LAMOST 光谱; MILES 光谱

中图分类号: TP29 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)10-3288-04

引言

根据现有的恒星演化理论, 恒星包括 Mg 元素在内的元素主要来源于短时标的 SN II 型超新星爆炸, 而 Fe 元素主要来源于更长时标的 SNIa 型超新星爆炸。因此, 恒星形成的时标越短, Mg 元素丰度[Mg/Fe]越高, 反之则 Mg 元素丰度[Mg/Fe]越低^[1]。因此星系内恒星的 Mg 元素丰度的分布可以揭示该星系的形成和演化历史。银河系是目前唯一的人类可以获得大量恒星 Mg 元素丰度信息的星系, 对银河系的恒星 Mg 元素丰度分布的研究将帮助我们深入了解银河系的形成和演化历史。

我国自主研发的大口径大视场望远镜郭守敬望远镜(LAMOST)目前获取了超过 800 万条中低分辨率的恒星光谱。获取这些恒星光谱的元素丰度, 将极大的扩展恒星元素丰度分布的信息, 帮助科学家更加深入的了解银河系的形成和演化历史。因此, 我们有必要研究如何从类似于 LAMOST 光谱的中低分辨率光谱中准确提取元素丰度。

目前关于从中低分辨率光谱中提取元素丰度研究较少, 主要使用模板匹配方法。Milone, Sansom 和 Sanchez-BlazquezLee 等研究了 MILES 光谱的 Mg 元素丰度提取方法, 主要通过模板匹配以及与高分辨光谱交叉的方法确定了 Mg 元素

丰度, 该方法确定的 Mg 元素丰度精度约为 0.12 dex^[2]。Xing 等使用模板匹配法确定 SLOAN 光谱中 Mg 元素丰度并从贫金属星中搜寻贫 Mg 丰度的恒星^[3]。Xing 等研究了使用模板匹配法确定低分辨率的 LAMOST 光谱中的 Mg 元素丰度并进而搜寻银晕中的贫 α 星^[4]。

以上研究主要使用模板匹配方法。但众所周知, 模板匹配法优化权重的方法复杂, 且易受噪声影响。本文将研究使用极限学习机算法(ELM)估计 Mg 元素丰度的方法。ELM 算法是 Huang 等提出的一种特殊的单隐层前向神经网络(SLFNs)^[5]。ELM 算法通过随机选择初始权重来提高训练速度, 同时避免传统神经网络可能陷入局部最优解的缺点。ELM 算法已经被应用于基因分析^[6], 图像分类^[7]等领域。本文将使用 ELM 算法确定中低分辨率的 MILES 光谱和 LAMOST 光谱的 Mg 丰度。实验结果显示该方法具有较高的准确度, 能够用于更大规模的 LAMOST 光谱的 Mg 丰度确定与特殊 Mg 丰度恒星搜寻。

1 ELM 算法简介

设 (x_i, t_i) ($i = 1, \dots, N$) 为样本数据, 其中 $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ 为数据, $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T$ 为对应于 x_i 的标签(目标值)。SLFNs 的数学模型为

收稿日期: 2017-07-10, 修订日期: 2018-02-19

基金项目: 国家自然科学基金项目(11603012, 11873037, 11603014, U1931209), 山东省青年基金项目(ZR2015AQ011), 中国博士后基金面上项目(2015M571124) 和山东大学青年学者未来计划项目(2016WHWLJH09)资助

作者简介: 卜育德, 1981 年生, 山东大学(威海)数学与统计学院副教授 e-mail: buyude001@163.com * 通讯联系人

$$\sum_{i=1}^{N_0} \beta_i g_i(x_j) = \sum_{i=1}^{N_0} \beta_i g(\omega_i x_j + b_i) = o_j$$

这里 $\omega_i = [\omega_{i1}, \omega_{i2}, \dots, \omega_{im}]^T$ 为连接输入层与隐藏层的第 i 个节点的权重 $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ 为连接第 i 个隐含节点与输出层之间的权重, b_i 为第 i 个隐含节点的偏置项, o_j 为 SLFNs 的第 j 个输出。我们的目标是根据训练数据确定 β_i , ω_i 和 b_i 以最小化下面的目标函数

$$E = \sum_{j=1}^N \left(\sum_{i=1}^{N_0} \beta_i g(\omega_i x_j + b_i) - t_j \right)^2$$

传统方法是通过梯度下降法来学习 ω_i , β_i 和 b_i , 而在 ELM 算法中我们首先随机设置 ω_i 和 b_i , 然后使用下述公式计算 β

$$\beta = H^{-1} T$$

这里 H^{-1} 为下述 H 矩阵的 Moore-Penrose 广义逆矩阵 H :

$$H = \begin{pmatrix} g(\omega_{N_0} x_1 + b_1) & \cdots & g(\omega_{N_0} x_1 + b_{N_0}) \\ \vdots & \ddots & \vdots \\ g(\omega_1 x_N + b_1) & \cdots & g(\omega_1 x_N + b_{N_0}) \end{pmatrix}_{N \times N_0}$$

$$T = [t_1^T, t_2^T, \dots, t_N^T]^T$$

由于 ELM 不需要确定最优的 ω_i 和 b_i , 因此 ELM 算法比传统的 SLFNs 方法效率高。而研究证实, 随机设置 ω_i 和 b_i 并不会降低算法的准确度。

2 实验部分

本文使用的数据包括两部分: MILES 光谱和 LAMOST 光谱。

我们使用的第一组数据包含有 752 个恒星光谱, 光谱来自于 MILES 最新释放的光谱库 MILES v9.1^[8]。MILES 恒星光谱库共包含有 985 个波长覆盖范围为 3 525~7 500 Å 分辨率为 2.3 Å(FWHM)的光谱, 但其中只有 752 个光谱具有 Mg 元素丰度, 其 Mg 元素丰度由 Milone 等通过模板匹配以及与高分辨光谱交叉的方法确定^[2]。第一组恒星数据的 [Mg/Fe] vs. [Fe/H] 分布图由图 1 给出。

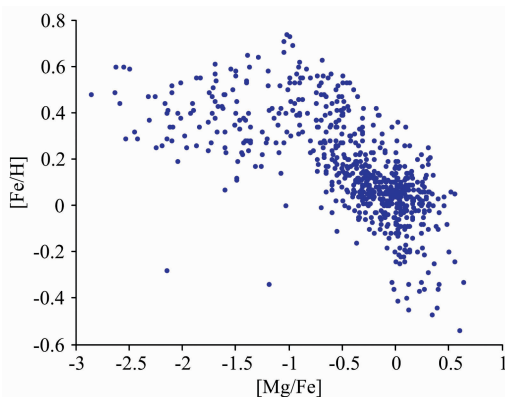


图 1 MILES 光谱的 [Mg/Fe] 与 [Fe/H] 分布图

Fig. 1 Distribution of MILES spectra in the [Mg/Fe]-[Fe/H] plane

实验中使用的第二组数据为 LAMOST 光谱, 共包含 13 536 条光谱, 其中信噪比 S/N 大于 50 的有 9 287 条。由于

LAMOST 释放的星表中没有提供 Mg 元素丰度, 我们使用的光谱的 Mg 元素丰度通过与 APOGEE 光谱库提供的 Mg 元素丰度交叉获得。APOGEE 是 SDSS III 项目的子项目^[9], 主要科学目标是使用高分辨 ($R \sim 20\,000$)、高信噪比 ($S/N \sim 100$) 的红外光谱给出超过 10 万颗红巨星的元素丰度以研究银河系的恒星形成过程和演化历史。这一组数据的 [Mg/Fe] vs. [Fe/H] 分布图由图 2 给出。

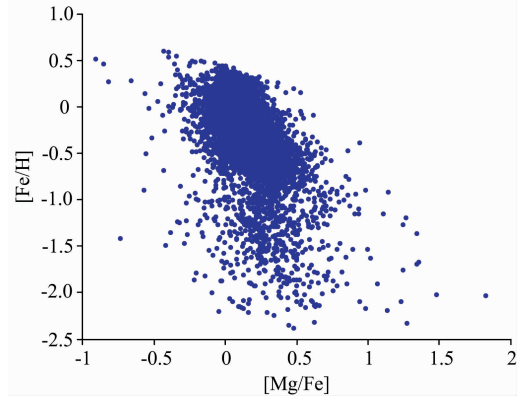


图 2 LAMOST 光谱的 [Mg/Fe] 与 [Fe/H] 分布图

Fig. 2 Distribution of LAMOST spectra in the [Mg/Fe]-[Fe/H] plane

实验中所有的数据将会被随机取出 20% 用来训练, 其余部分用来测试。而实验结果主要由如下两个标准衡量: Δ , 目标值与预测值之间的差距的中位数; σ , 目标值与预测值之间的差距的标准差。

3 结果与讨论

3.1 Mg 元素丰度估计

我们首先在 MILES 光谱上使用 ELM 算法进行 Mg 丰度估计, 实验结果由图 3 给出。结果显示, ELM 算法给出的精度为 $\Delta = 0.009\,9\text{ dex}$ ($\sigma = 0.15\text{ dex}$), 同 MILES 提供的 Mg 元素丰度精度类似。因此, ELM 算法能够较为精确的确定 MILES 光谱的 Mg 元素的丰度。

我们又使用 LAMOST 的信噪比 $S/N > 50$ 的 9 287 条光谱进行了实验, 实验结果由图 4 给出。结果显示, ELM 算法给出的精度为 $\Delta = 0.002\,7\text{ dex}$ ($\sigma = 0.11\text{ dex}$)。该结果比 MILES 光谱上的结果要好, 原因可能是 MILES 光谱提供 Mg 元素丰度本身精度不高, 导致无法精确的训练算法。

3.2 信噪比对结果的影响

本实验中我们使用 LAMOST 光谱进行实验。把 LAMOST 光谱按照信噪比分成如下 7 组数据: D1: $S/N \geq 100$; D2: $80 \leq S/N < 100$; D3: $60 \leq S/N < 80$; D4: $40 \leq S/N < 60$; D5: $20 \leq S/N < 40$; D6: $10 \leq S/N < 20$; D7: $S/N < 10$ 。然后对每一组数据随机选取 20% 作为训练数据, 其余部分用来测试。实验结果由表 1 给出。

结果显示, 随着信噪比降低, 估计结果精度会逐渐下降。在数据集 D4 上的精度为 $\Delta = -0.005\,7\text{ dex}$ ($\sigma = 0.14$)

dex), 与 ELM 算法在 MILES 光谱上的精度类似。而在信噪比小于 40 的数据集 D5, D6 和 D7 上估计精度则较差。因此, 在使用 ELM 算法时, 建议使用具有较高信噪比的 LAMOST 光谱。考虑到 LAMOSTDR4 中包含的信噪比大于 40 的光谱有 1 871 952 条, 而我们的算法可以应用于这些海量的光谱中, 因此有很重要的应用价值。

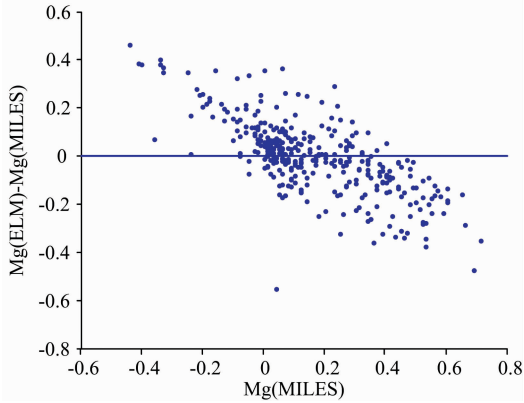


图 3 ELM 给出的 MILES 光谱的 Mg 丰度估计结果

x 轴为 MILES 给出的 Mg 丰度估计, y 轴为 ELM 算法给出的结果与 MILES 给出的结果之差

Fig. 3 Mg abundance of MILES spectra determined with ELM algorithm

The x -axis represents the Mg abundance from MILES catalog, and y -axis represents the difference between results determined with ELM with those from MILES catalog

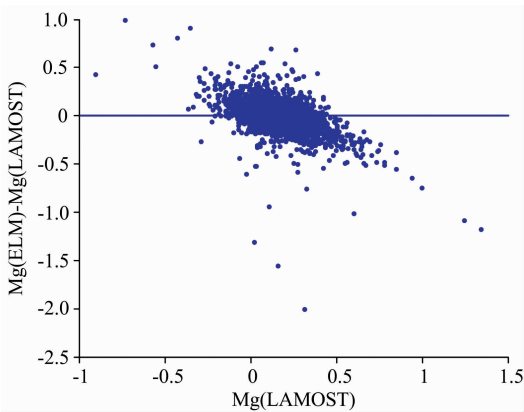


图 4 ELM 给出的 LAMOST 光谱的 Mg 丰度估计

x 轴为 LAMOST 给出的 Mg 丰度, y 轴为 ELM 给出的 Mg 丰度与 LAMOST 光谱的 Mg 丰度之差

Fig. 4 Mg abundance of the LAMOST spectra determined with ELM algorithm

The x -axis represents the Mg abundance from LAMOST catalog, and y -axis represents the difference between results determined with ELM with those from LAMOST catalog

3.3 数据预处理对结果的影响

上述实验中我们均直接使用了 LAMOST 光谱数据, 没有对光谱数据做降维处理。一般情形下, 使用 PCA 对数据做预处理, 将会减少光谱的噪声, 提高算法的效率。为检验

PCA 能否提升 ELM 算法的结果, 我们用 PCA 降维之后的数据作为 ELM 算法的输入对 Mg 丰度进行了估计, 并与使用未降维的光谱的结果进行了对比, 结果由表 2 给出。

表 1 使用不同信噪比的光谱的 Mg 元素丰度估计结果
Table 1 Performance comparison of different algorithms on estimating Ma abundances

数据集	光谱数	$\Delta(\sigma)$
D1	4 610	0.005 9(0.086)
D2	1 477	0.005 5(0.10)
D3	2 010	0.009 4(0.12)
D4	2 551	-0.005 7(0.14)
D5	2 234	0.012(0.17)
D6	567	0.040(0.25)
D7	87	0.049(0.31)

表 2 使用 PC 与使用光谱的结果比较

Table 2 Comparison between results from PC those from the spectra

数据集	PC	光谱
D1	0.004 7(0.090)	0.005 9(0.086)
D2	0.015(0.11)	0.005 5(0.10)
D3	0.005 4(0.13)	0.009 4(0.12)
D4	0.002 1(0.16)	-0.005 7(0.14)
D5	-0.005 5(0.18)	0.012(0.17)
D6	-0.003 4(0.27)	0.040(0.25)
D7	-0.004 3(0.34)	0.049(0.31)

结果显示, 使用 PCA 预处理过的数据得到的丰度估计结果与使用光谱得到的结果精度类似。这说明, 使用 ELM 算法估计 Mg 元素丰度不需要光谱预处理即可得到较为精确的估计结果, 这说明了 ELM 算法能够从光谱提供的信息中准确提取出和 Mg 丰度有关的信息而不需要借助于其他数据预处理。

3.4 不同算法结果的比较

为了进一步展示 ELM 算法的精度, 我们把 ELM 算法同如下常用的算法进行了对比: 支持向量回归(SVR), 线性回归(LR), 高斯过程回归(GPR), 核回归(KR), 人工神经网络(ANNs)。我们使用的数据为信噪比大于 50 的 LAMOST 数据, 结果由表 3 给出

表 3 不同算法结果对比

Table 3 Comparison of results of different algorithms

算法	Δ	σ
ELM	0.002 7	0.11
GPR	0.003 4	0.12
SVR	0.001 2	0.13
KR	0.014	0.12
LR	0.007 7	0.51
ANN	-0.045	0.23

结果显示, 相比于其他算法, ELM 算法给出的估计精度最高。因此, ELM 算法可以用来估计 Mg 元素的丰度。

4 结 论

研究了一种新的估计恒星元素丰度的方法: ELM 方法。该方法基于单层神经网络方法, 但由于采用了随机设置某些参数的方法, 极大提高了运算速度, 同时不影响运算精度。我们把该方法应用于估计 MILES 光谱与 LAMOST 光谱的 Mg 丰度。实验结果显示:

(1) ELM 估计 MILES 光谱的 Mg 元素丰度精度为 $\Delta = 0.0099 \text{ dex}$ ($\sigma = 0.15 \text{ dex}$), 估计信噪比 $S/N > 50$ 的 LAMOST 光谱的 Mg 丰度的精度为 $\Delta = 0.0027 \text{ dex}$ ($\sigma = 0.11$

dex)。这说明 ELM 估计恒星的 Mg 元素丰度精度较高, 可以被用于中低分辨率光谱的 Mg 元素丰度的确定。

(2) ELM 算法的结果会随着信噪比的下降而精度降低。其在信噪比 $S/N < 40$ 的光谱上结果低于 $\Delta = 0.012 \text{ dex}$ ($\sigma = 0.17 \text{ dex}$)。因此, 在实际应用中应把 ELM 算法用于信噪比大于 40 的光谱。

(3) 使用 PCA 对数据进行预处理并不能提高 ELM 算法结果的精度。这说明 ELM 算法本身能够从光谱中提取到和 Mg 元素丰度相关的特征, 具有很强的特征提取能力。

(4) ELM 算法同 GPR 算法、SVR 算法、KR 算法、LR 算法和 ANN 算法的比较结果显示, ELM 算法的精度最高。因此 ELM 算法是一种可靠的 Mg 丰度估计方法, 可以应用于 LAMOST 后续光谱的 Mg 丰度估计。

References

- [1] Lee Y S, Beers T C, Prieto C A, et al. *The Astronomical Journal*, 2010, 141(3): 90.
- [2] Milone A D C, Sansom A E, Sanchez-Blazquez P. *Monthly Notices of the Royal Astronomical Society*, 2011, 414(2): 1227.
- [3] Xing Qianfan, Zhao Gang. *The Astrophysical Journal*, 2014, 790(1): 33.
- [4] Xing Qianfan, Zhao Gang, Zhang Yong, et al. *Research in Astronomy and Astrophysics*, 2015, 15(8): 1275.
- [5] Huang G B, Zhu Q Y, Siew C K. *Neurocomputing*, 2006, 70(1): 489.
- [6] Savojardo C, Fariselli P, Casadio R. *Bioinformatics*, 2011, 27(16): 2224.
- [7] Akusok A, Bjork K, Miche Y, et al. *IEEE Open Access*, 2015, 3: 1011.
- [8] Falcon-Barroso J, Sanchez-Blazquez P, Vazdekis A, et al. *Astronomy & Astrophysics*, 2011, 532: A95.
- [9] Prieto C A, Majewski S R, Schiavon R, et al. *Astronomische Nachrichten*, 2008, 329: 1018.

Estimation of Mg Abundances from Low Resolution Spectra Using ELM Algorithm

BU Yu-de^{1*}, PAN Jing-chang², YI Zhen-ping²

1. School of Mathematics and Statistics, Shandong University, Weihai, Weihai 264209, China

2. School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai 264209, China

Abstract In this paper, we mainly study a new method for estimating the Mg abundance of stars from the medium and low resolution spectra based on ELM algorithm. LAMOST provides us with a massive low-resolution spectrum, and determining the abundance of Mg elements in these spectra will help us understand the history and evolution of the Milky Way. At present, the traditional method of determining the abundance of Mg element from medium and low resolution spectra is the template matching method. However, this method is difficult to optimize parameters and is sensitive to noise. Therefore, it is necessary to study new methods to estimate the Mg abundance. The experiment show that ELM algorithm is a good alternative to traditional method. The accuracy of ELM algorithm on MILES spectra is 0.0099 (0.15) dex, while on the LAMOST spectra with signal-to-noise ratios larger than 50 it is 0.0027 (0.11) dex. A comparison of ELM with other four algorithms shows that ELM algorithm can accurately estimate the abundance of Mg elements from low resolution spectra and can be applied to the LAMOST spectra.

Keywords ELM algorithm; Abundance; LAMOST spectra; MILES spectra

(Received Jul. 10, 2017; accepted Feb. 19, 2018)

* Corresponding author