ELSEVIER

CrossMark

Chinese Astronomy and Astrophysics 39 (2015) 319–329

CHINESE
ASTRONOMY
AND ASTROPHYSICS

Application of Multi-task Lasso Regression in the Parametrization of Stellar Spectra† ⋆

CHANG Li-na   ZHANG Pei-ai△

College of Information Science and Technology, Jinan University, Guangzhou 510632

Abstract The multi-task learning approaches have attracted the increasing attention in the fields of machine learning, computer vision, and artificial intelligence. By utilizing the correlations in tasks, learning multiple related tasks simultaneously is better than learning each task independently. An efficient multi-task Lasso (Least Absolute Shrinkage Selection and Operator) regression algorithm is proposed in this paper to estimate the physical parameters of stellar spectra. It not only can obtain the information about the common features of the different physical parameters, but also can preserve effectively their own peculiar features. Experiments were done based on the ELODIE synthetic spectral data simulated with the stellar atmospheric model, and on the SDSS data released by the American large-scale survey Sloan. The estimation precision of our model is better than those of the methods in the related literature, especially for the estimates of the gravitational acceleration ($\lg g$) and the chemical abundance ([Fe/H]). In the experiments we changed the spectral resolution, and applied the noises with different signal-to-noise ratios (SNRs) to the spectral data, so as to illustrate the stability of the model. The results show that the model is influenced by both the resolution and the noise. But the influence of the noise is larger than that of the resolution. In general, the multi-task Lasso regression algorithm is easy to operate, it has a strong stability, and can also improve the overall prediction accuracy of the model.

Key words stars: fundamental parameters—methods: data analysis—methods: statistical—methods: miscellaneous

† Supported by the General Project of the Ministry of Education for Researches of Humane and Social Sciences (11YJAZH118)

Received 2014–05–16; revised version 2014–08–06

⋆ A translation of *Acta Astron. Sin.* Vol. 56, No. 1, pp. 26–34, 2015

△ qzhzhang@163.com

0275-1062/15/$-see front matter
doi:10.1016/j.chinastron.2015.07.004

1. INTRODUCTION

At present the large-aperture and large-field telescopes, such as the Chinese LAMOST[1−2], the American Sloan digital survey telescope[3−4], etc., can obtain a large number of spectral data. How to measure the physical parameters of stellar spectra in the numerous quantity of spectral data becomes the most fundamental and important subject in the spectral data processing of celestial bodies. The physical parameters of stellar spectra are mainly the surface effective temperature (T_{eff}), the gravitational acceleration ($\lg g$), and the chemical abundance ([Fe/H]). So far many automated parameter estimation methods of stellar spectra have been proposed, they are mainly the artificial neural network (ANN) algorithm and the minimum distance method (MDM). For example, Baikler-Jones[5] designed an ANN of 820:5:5:1 structure for the prediction of the stellar effective surface temperature, and further developed an ANN system of feedforward multi-layer perception with two "hidden layers" in 2000. And the K-nearest-neighbor algorithm of Fuentes et al.[6], the weighted average algorithm of Allende[7], and the non-parametric regression of varying window-width of Zhang et al.[8−9] are all the varieties of the MDM.

Because of the numerous quantity of spectral data, and the various noise interferences in the transmission and receiving processes of spectra, such as the photon noise, sky line, and instrumental noise, the finally estimated results of stellar physical parameters will be influenced. Before a model prediction is made, the dimension reduction and noise elimination should be made for the high-dimensional data, the related methods are the principal components analysis (PCA), filtering method, and so on. Among them, the Lasso (least absolute shrinkage selection and operator) algorithm proposed by Tibshirani[10] in 1996 is more attractive, because of its rapid calculation and high accuracy. The stellar spectra contain the information of stellar physical parameters, but at present many models deal with the stellar physical parameters separately, so that the potential relations among the different physical parameters are neglected. In recent years, in the fields of machine learning, computer vision, and artificial intelligence, the multi-task learning (MTL) method has attracted the attention from many researchers, which is purposed to acquire the potential relations of different tasks, to learn multiple related tasks simultaneously, and therefore to utilize the rich information among the different tasks. This learning method is favorable for the mutual learning of tasks, and can exhibit distinctly its advantage for improving the prediction effectiveness and generalization performance of the prediction model. For example, Evgenious et al.[11] applied the method of multi-task support vector machine to improving the prediction accuracy of consumers' preference; by experiments, Bakker et al.[12] demonstrated that in the condition of small number of image categories, the multi-task classification can improve the generalization performance of the classifier. Although the Lasso algorithm can make the every task learn effectively and independently[13], but it may neglect the potential relations among the tasks, causing the overfitting of the data, and affecting the final performance of the model. In the multi-task learning, by the effective minimization of the L_{21}-norm form,