第42卷，第6期
2022年6月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 42, No. 6, pp1881-1885
June, 2022

基于FFCNN的二维恒星光谱分类

逯亚坤[1]，邱　波[1*]，罗阿理[2]，郭小雨[1]，王林倩[1]，曹冠龙[1]，白仲瑞[2]，陈建军[2]

1. 河北工业大学，天津　300400
2. 中国科学院国家天文台，北京　100012

摘　要　天体光谱处理中的一项基本任务是对大量的恒星光谱进行自动分类。到目前为止，恒星光谱的分类工作多是基于一维光谱数据。该研究打破传统的天体光谱数据处理流程，提出了基于二维恒星光谱分类的方法。在LAMOST(the large sky area multi-object fiber spectroscopic telescope)的数据处理流程中，所有的一维光谱都是由二维光谱抽谱、合并得来。二维光谱是由光谱仪产生的图像，包括蓝端图像和红端图像。基于LAMOST二维光谱数据，提出了特征融合卷积神经网络(FFCNN)分类模型，用于二维恒星光谱的分类。该模型是一个有监督的算法，通过两个CNN模型分别提取蓝端图像和红端图像的特征，然后将二者进行融合得到新的特征，再利用CNN对新特征进行分类。所使用的数据全部来源于LAMOST，我们在LMOST DR7中随机选择了一批源，然后获得了它们的二维光谱。一共有14 840根F，G和K型恒星的二维光谱用于FFCNN模型的训练，其中包括7 420根蓝端光谱和7 420根红端光谱。由于三类恒星光谱的数量并不均衡，在训练的过程中分别为每类恒星光谱设置了不同权重，防止模型出现分类失衡现象。同时，为了加快模型收敛，对二维光谱数据采用Z-score归一化处理。此外，为了充分利用所有样本，提高模型的可靠度，采用五折交叉验证的方法验证模型。3 710根二维光谱用作测试集，使用准确率、精确率、召回率和F1-score来对FFCNN模型的性能进行评价。实验结果显示，F，G和K型恒星的精确率分别达到87.6%，79.2%和88.5%，而且它们超过了一维光谱分类的结果。实验结果证明基于FFCNN的二维恒星光谱分类是一种有效的方法，它也为恒星光谱的处理提供了新的思路和方法。

关键词　二维恒星光谱；光谱分类；FFCNN模型；归一化；交叉验证

中图分类号：P157.2　　**文献标识码**：A　　**DOI**：10.3964/j.issn.1000-0593(2022)06-1881-05

引　言

现代技术的发展极大地改善了天文观测能力。随着大量巡天项目的运行，例如英国和澳大利亚的2dF项目，美国的SDSS项目，欧洲的Gaia项目[1]和中国的LAMOST项目[2]等，天文数据呈指数增长。恒星光谱分类是天文数据分析的基本任务之一。传统的恒星光谱分类方法是基于Morgan-Kenan系统[3](MK系统)。根据MK分类系统，天文学家按照温度由高到低，将恒星分为七个类别：O，B，A，F，G，K和M型。每种类型又可以细分为从0到9共10个子类型。MK分类系统主要是用待测光谱与标准星型模板光谱进行比较[4]。它效率低下，不适合海量数据的处理。

最近几年，研究人员将机器学习理论应用到天文学领域并提出很多光谱分类算法[5-7]。但是机器学习算法存在很多问题，例如特征提取能力有限，使用场景有限，泛化能力差等。这些缺陷导致恒星光谱分类精度较低。深度学习是机器学习的发展，因为其强大的特征提取能力，已经在很多领域都取得成功[8]。

Liu[9]等提出了1D SSCNN模型对F，G和K型恒星的光谱进行分类。该模型对标签使用one-hot编码，从而提高模型的学习能力。实验结果表明，一维SSCNN的分类准确率达到90%，其效果优于RF，KNN和SVM算法。由于天体中O型恒星很难被观察，所以，与其他恒星光谱相比，LAMOST光谱库中O型恒星的数量相对较少。数据的平衡对于深度学习的分类性能有重要影响。为了解决恒星光谱分

收稿日期：2021-05-02，**修订日期**：2021-06-20

基金项目：国家自然科学基金委员会-中国科学院天文联合基金项目(U1931134)，河北省自然科学基金项目(A2020202001)，中国科学院天文大科学研究中心LAMOST重大成果培育项目资助

作者简介：逯亚坤，1995年生，河北工业大学硕士研究生　　e-mail：846296206@qq.com

*通讯作者　　e-mail：qiubo@hebut.edu.cn

布不平衡的问题，Zheng 等[10]提出了由 SGAN 和 CNN 组成的半监督分类模型，该模型通过 SGAN 生成少量的 O 型光谱，来达到数据的平衡，防止在训练过程中出现过拟合现象。对于深度学习来说，网络结构越深，特征提取能力越强。但过深的网络结构会导致梯度消失和梯度爆炸。为了解决这个问题，Zou 等[11]将 RAC-Net 网络应用到恒星光谱分类任务上。该模型可以使用残差模块和注意力机制来增加模型的深度和模型重要通道的关注度。

尽管这些基于 1D 光谱的光谱分类方法已经可以取得很高的准确率，但是他们需要高信噪比的 1D 光谱。LAMOST 中的 1D 光谱都是经过 2D 光谱抽谱、合并得来。迄今为止，天文学家对光谱的分析都是基于 1D 光谱数据[12]。很少有学者利用 2D 光谱信息来研究天文学的基本任务。从数据的角度来看，分布在 2D 空间中的数据具有更多的特征，例如空间特征和纹理特征。将 2D 光谱变换为 1D 光谱，光谱的 2D 信息会丢失。而且如果 2D 光谱的信噪比很低，便无法对其进行有效的抽谱，也无法对其进行有效的分析。LAMOST 大概有 1/5 的光谱因为这些原因无法被使用。

因此，我们尝试改变仅基于 1D 光谱的分析方法，提出了基于特征融合的卷积神经网络用于 2D 恒星光谱的分类。同时与 1D 恒星光谱分类算法对比。

1 LAMOST 2D 光谱和数据归一化

1.1 LAMOST 2D 光谱

LAMOST 是我国口径最大的望远镜，配备了 16 个光谱仪和 32 台 CCD 相机。每台 CCD 摄像机分配了 250 根光纤，并同时对光纤的蓝端和红端成像。每根光纤呈带状分布，如图 1 所示。

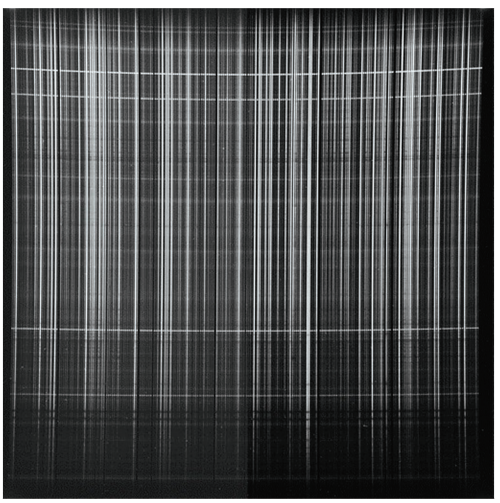

图 1 LAMOST 目标光谱图像

Fig. 1 Spectral image of LAMOST object

LAMOST 的每张 2D 光谱图可以记录 250 根光纤目标。光谱密集地排列在 2D 光谱图像中，相邻的两个光谱之间会存在交叉污染。从中心轨迹开始，越远的地方污染越大。图 2 是一根 2D 光谱图像，其中长度为 100 像素，宽度为 15 像素。从图中可以看到，中间的 9 个像素更亮，分布在两边的像素比较暗。为了防止光谱之间的污染，我们只选择中心轨迹周围的 9 个像素作为其有效数据。

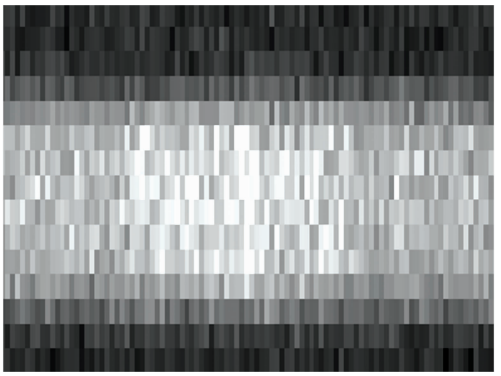

图 2 一根 2D 光谱图像

Fig. 2 An image of 2D spectrum

1.2 数据归一化

数据标准化是处理数据时的基本操作。不同光谱的最大亮度差异很大，可能会影响数据的分类结果。因此，为了要消除不同尺度的影响，需要对 2D 光谱归一化。本文使用的归一化方法为 Z-score。经过其处理后的数据分布符合标准正态分布，即平均值为 0，标准偏差为 1。转化公式为

$$x^* = \frac{x-\mu}{\sigma} \tag{1}$$

式(1)中，x^* 为转换后的数据，x 为原始数据，μ 为原始数据的均值，σ 为原始数据的方差。

2 FFCNN 算法介绍

2.1 CNN

CNN 是目前最流行的算法，它主要由输入层、隐藏层和输出层组成，拥有强大的提取图像特征的能力[13-14]。输入层和输出层都是单层结构，隐藏层通常都是多层。每一层的节点叫作神经元，其输入输出关系可以表示为

$$Y_{\text{out}} = f\left(b + \sum_{i=1}^{N} w_i X_i\right) \tag{2}$$

式(2)中，Y_{out} 为神经元的输出，f 为激活函数，b 为偏置，w 为权重，X 为神经元的输入。

2.2 FFCNN

本文基于传统的 CNN 模型，提出了 FFCNN 模型。LAMOST 的 2D 光谱包括蓝端和红端数据，蓝端波长范围大概为 3 700～5 900 Å，红端波长范围大概为 5 700～9 000 Å。不同的恒星光谱在不同的波长处具有不同的流量，仅使用蓝端或红端不足以对所有类型的恒星光谱进行分类。传统的光谱处理流程，先分别对 2D 光谱的蓝、红端进行抽谱操作，然后再将两端数据进行合并，该方法属于数据级的特征融合。本文中，我们提出了特征级的融合方法，直接使用两个 CNN 结构，分别提取 2D 光谱的蓝端和红端特征，然后将提取出的特征进行融合，得到新的特征，CNN 将继续学习这些特征，从而对光谱分类。

FFCNN 的网络结构如图 3 所示。FFCNN 主要包含三个模块：卷积模块、特征融合模块和全连接模块。在卷积模块中，它包含两个卷积层，一个 BN 层和一个最大池化层。每个卷积层使用 Relu 激活函数，并且使用 Dropout 技术防止模型过拟合。特征融合模块包含一个拼接层，其功能是融合从两个 CNN 分支中提取出的特征。全连接层模块包含两个全连接层。第一个全连接的节点数为 1 024，使用 Relu 激活函数。第二个全连接层的节点数为 3，使用 softmax 激活函数。

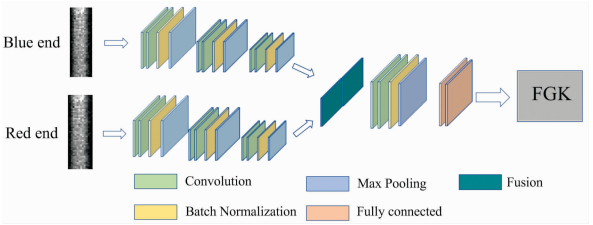

图 3　FFCNN 模型的网络结构

Fig. 3　The structure of FFCNN

3　实验结果与讨论

3.1　评价指标

为了验证 FFCNN 和 2D 光谱在恒星光谱分类中的可行性，使用准确率、精确率、召回率和 F1-score 来评估模型。准确率代表所有预测正确的光谱占测试集中光谱的比例。精确率代表正确预测的光谱与所有预测为该类光谱的比例。召回率代表正确预测的光谱占测试集中所有光谱的比例。F1 得分是精确度和召回率的调和平均值，可以全面的表示模型的性能。

3.2　2D 光谱实验和结果分析

实验环境为 Intel Core i5 CPU，8G RAM，Windows 10 和 Python 3.7。实验中的所有数据均来自 LAMOST DR7。

一共 14 840 根光谱，包括 7 420 根蓝端光谱和 7 420 根红端光谱用于 FFCNN 模型的训练。训练集中 F，G 和 K 数量之比约为 2∶3∶6。为了缓解数据不均衡对模型训练的影响，在 FFCNN 训练期间，分别为 F、G 和 K 设置 3∶2∶1 的权重。训练时，权重值越大，越能受到模型的关注。

与传统图像不同，经过处理后的 2D 光谱在波长方向上有 4 136 个像素，在空间方向上有 9 个像素，横纵比很大。在传统 CNN 模型中，卷积核大小一般为 3×3 或 5×5，但它们显然是不适合 2D 光谱。为了更好地提取 2D 光谱的波长方向和空间方向的特征，我们使用 3×25 的卷积核。空间方向的长度为 9 个像素，因此我们在这个方向上的卷积核长度设置为 3，这是流行的卷积核大小。波长方向上的长度为 4 196 像素。为了提高 CNN 的感受野，我们在这个方向上卷积核长度设置为 25。更长的卷积核将花费更多的训练时间，而且不能提高模型的准确率。该方向的卷积核长度与训练时间和模型准确率的关系，如图 4 所示。

为了充分利用所有样本，并提高模型的可靠性，我们使用 5 折交叉验证来验证模型的性能。即数据被平均分为 5 份，4 份被用于训练，1 份被用于测试。一共进行了 5 次试验，然后获得平均准确率。

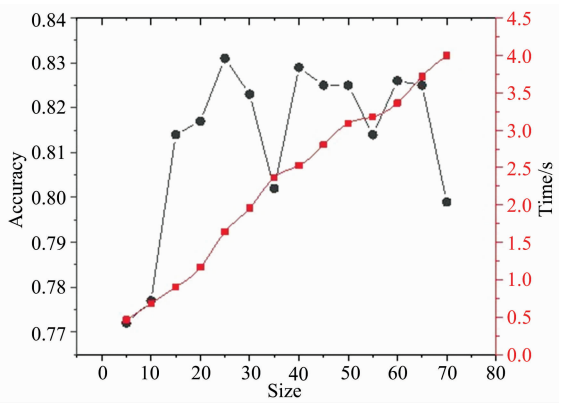

图 4　在波长方向，不同长度卷积核对模型的准确率和训练时间的影响

Fig. 4　The influence of convolution kernel with different length on the accuracy and the time spent in the model training

一共使用了 3 710 根光谱，包括 1 855 根蓝端和 1 855 根红端光谱，测试模型的性能。2D 光谱分类实验结果，如图 5 所示。

图 5 是一个混淆矩阵，图中显示大多数测试数据分布在其主对角线上，也就是说，这些光谱可以被正确分类。

FFCNN 模型的准确率、精确率、召回率和 F1-score 如表 1 所示。F，G 和 K 型恒星的精确率分别为 87.6%，79.2%和 88.5%，说明 FFCNN 可以准确地区分 2D 恒星光谱的类别。

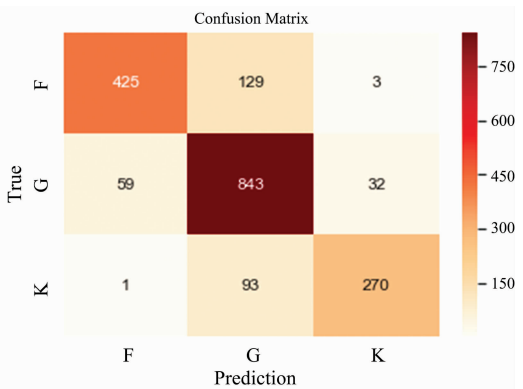

图 5　测试集中所有样本的分类结果

Fig. 5　Classification results of all samples in the testing set

表 1　FFCNN 模型的精确率、召回率、F1-score 和准确率

Table 1　The precision, recall, F1-score and the accuracy of FFCNN model

	F	G	K
Precision	0.876	0.792	0.885
Recall	0.763	0.903	0.742
F1-score	0.816	0.843	0.807
Accuracy		0.829	

3.3 2D 和 1D 光谱分类结果的比较

为了更好地研究 2D 光谱在恒星光谱分类上的可行性，本实验还比较了 1D 光谱分类的结果。

使用孔径抽谱[15]的方法，将 2D 光谱转换为 1D 光谱，并将蓝端和红端光谱拼接，使用 Liu 等[9]提出的 1D SSCNN 方法对 1D 光谱进行分类。对比实验结果如图 6 所示。F，G 和 K 型恒星在 1D 光谱上的精确度分别为 73.7%，79.1%和 69.8%。只有 G 型恒星在 1D 光谱上和 2D 光谱上的分类结果相近，对于 F 和 K 型恒星来说，2D 光谱的分类精确度是明显高于 1D 光谱的，其精确度分别提高了 14%和 19%。

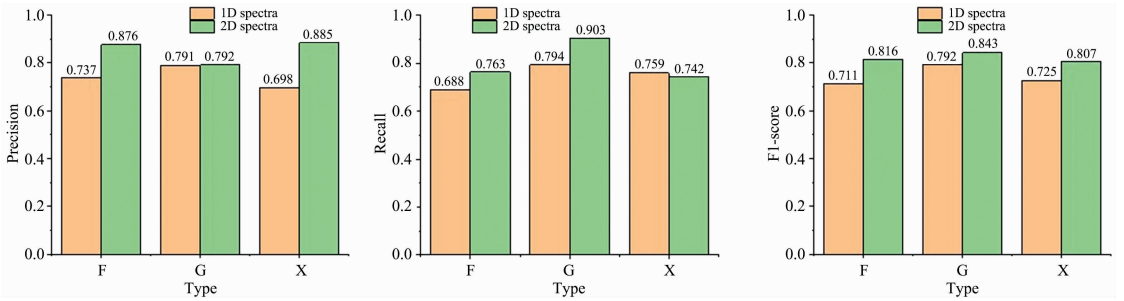

图 6 2D 光谱分类和 1D 光谱分类结果的比较

Fig. 6 The comparison of classification results of using 2D and 1D spectra

4 结 论

用一种新的方法来处理天体光谱，以克服当前处理 1D 光谱的瓶颈。本文提出了 FFCNN 分类模型用于 LAMOST 2D 恒星光谱的自动分类。与传统的分析光谱的流程不同，该方法不再需要对 2D 光谱进行抽谱和合并，而是将 FFCNN 直接应用于 2D 光谱。在 FFCNN 中，2D 光谱被划分为训练集和测试集。测试集中包含 14 840 根光谱用于模型的训练，训练集中包含 3 710 根光谱用于模型的评估。实验结果证明 FFCNN 网络可以有效提取光谱的 2D 特征，F，G 和 K 型恒星的精确率分别达到 87.6%，79.2%和 88.5%。该方法不仅能提供可靠的光谱分类结果，同时简化了 2D 光谱转换为 1D 光谱的复杂操作，提高了光谱的利用率。

References

[1] Chen P, Shan H, Gao Y. Astrophysics and Space Science, 2011, 331: 63.
[2] Yao S, Wu X, Li Y, et al. The Astrophysical Journal Supplement Series, 2019, 240: 6.
[3] Morgan W, Keenan P. Annual Review of Astronomy and Astrophysics, 1973, 11: 29.
[4] Gray R, Corbally C. The Astronomical Journal, 2013, 147: 80.
[5] Fabbro S, Venn K A, et al. Monthly Notices of the Royal Astronomical Society, 2018, 475: 2978.
[6] Hoyle B. Astronomy and Computing, 2016, 16: 34.
[7] Hon M, Stello D, Yu J. Monthly Notices of the Royal Astronomical Society, 2017, 469: 4578.
[8] Schmidhuber J. Neural Networks, 2015, 61: 85.
[9] Liu W, Zhu M, et al. Monthly Notices of the Royal Astronomical Society, 2019, 483: 4774.
[10] Zheng Z P, Qiu B, Luo A L. Publications of the Astronomical Society of the Pacific, 2020, 132: 024504.
[11] Zou Z, Zhu T, Xu L, et al. Publications of the Astronomical Society of the Pacific, 2020, 132: 044503.
[12] Li G, Zhang H, Bai Z. Publications of the Astronomical Society of the Pacific, 2015, 127: 552.
[13] He X, Chen Y. IEEE Geoscience and Remote Sensing Letters, 2021, 18(5): 876.
[14] Hang R, Li Z, Liu Q, et al. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(3): 2281.
[15] Ritter A, Hyde E A, Parker Q A. Publications of the Astronomical Society of the Pacific, 2014, 126: 170.

Classification of 2D Stellar Spectra Based on FFCNN

LU Ya-kun[1], QIU Bo[1*], LUO A-li[2], GUO Xiao-yu[1], WANG Lin-qian[1], CAO Guan-long[1], BAI Zhong-rui[2], CHEN Jian-jun[2]

1. Hebei University of Technology, Tianjin 300400, China
2. National Astronomical Observatory, Chinese Academy of Sciences, Beijing 100012, China

Abstract Automatic classification of many stellar spectra is a basic task in celestial spectral processing. So far, the classification of star spectra is based on one-dimensional (1D) spectra. This paper proposes a new method based on two-dimensional(2D) stellar spectral classification. In the data processing process of LAMOST (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope), 1D spectra are extracted and combined with 2D spectra, which are the images produced by a spectrometer, including blue end and red end. Based on LAMOST 2D spectra, a convolutional neural network (FFCNN) classification model is proposed for stellar spectral classification. The model is a supervised algorithm which extracts the features of the blue end and red end respectively through two CNN models. And the model fuses the two features to get new features and uses CNN to classify the new features. The data used in this work are all from LAMOST. A batch of sources are randomly selected in LAMOST DR 7, and their 2D spectra are obtained. There are 14 840 F, G, and K stars in 2D spectra for training the FFCNN model, including 7 420 blue end and 7 420 red end spectra. The number of three kinds of stellar spectra is not balanced. Different weights are set for each kind of stellar spectra in the training process to prevent the classification imbalance. At the same time, to accelerate the model's convergence, the Z-score normalization method is used for 2D spectra. In addition, five-fold cross-validation is used to improve the model's sample utilization and reliability. 3 710 2D spectra are used as the test set, and the accuracy, precision, recall and F1-score are used to evaluate the performance of the FFCNN model. Experimental results show that the precision of F, G, and K type stars reach 87.6%, 79.2%, and 88.5%, respectively, and they exceed the results of 1D spectral classification. The experimental results prove that the 2D stellar spectral classification based on FFCNN is an effective method, and it also provides new ideas and methods for the processing of stellar spectra.

Keywords Two-dimensional stellar spectra; Spectral classification; FFCNN model; Normalized; Cross-validation

(Received May 2, 2021; accepted Jun. 20, 2021)

* Corresponding author