Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **431,** 1800–1811 (2013)
Advance Access publication 2013 March 20

doi:10.1093/mnras/stt298

Mining unusual and rare stellar spectra from large spectroscopic survey data sets using the outlier-detection method

Peng Wei,[1,2] Ali Luo,[1,3]★ Yinbi Li,[1] Jingchang Pan,[3] Liangping Tu,[1,4] Bin Jiang,[1,2,3] Xiao Kong,[1] Zhixin Shi,[1,2,5] Zhenping Yi,[1,2,3] Fengfei Wang,[1,2] Jie Liu[3] and Yongheng Zhao[1]

[1]*Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China*
[2]*University of Chinese Academy of Sciences, Beijing 100049, China*
[3]*School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China*
[4]*School of Science, Liaoning University of Science and Technology, Anshan 144051, China*
[5]*Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*

Accepted 2013 February 14. Received 2013 February 14; in original form 2012 November 15

ABSTRACT

The large number of spectra obtained from sky surveys such as the Sloan Digital Sky Survey (SDSS) and the survey executed by the Large sky Area Multi-Object fibre Spectroscopic Telescope (LAMOST, also called GuoShouJing Telescope) provide us with opportunities to search for peculiar or even unknown types of spectra. In response to the limitations of existing methods, a novel outlier-mining method, the Monte Carlo Local Outlier Factor (MCLOF), is proposed in this paper, which can be used to highlight unusual and rare spectra from large spectroscopic survey data sets. The MCLOF method exposes outliers automatically and efficiently by marking each spectrum with a number, i.e. using *outlier index* as a flag for an unusual and rare spectrum. The Local Outlier Factor (LOF) represents how unusual and rare a spectrum is compared with other spectra and the Monte Carlo method is used to compute the global LOF for each spectrum by randomly selecting samples in each independent iteration. Our MCLOF method is applied to over half a million stellar spectra (classified as STAR by the SDSS Pipeline) from the SDSS data release 8 (DR8) and a total of 37 033 spectra are selected as outliers with signal-to-noise ratio (S/N) ≥ 3 and *outlier index* ≥ 0.85. Some of these outliers are shown to be binary stars, emission-line stars, carbon stars and stars with unusual continuum. The results show that our proposed method can efficiently highlight these unusual spectra from the survey data sets. In addition, some relatively rare and interesting spectra are selected, indicating that the proposed method can also be used to mine rare, even unknown, spectra. The proposed method can be applicable not only to spectral survey data sets but also to other types of survey data sets. The spectra of all peculiar objects selected by our MCLOF method are available from a user-friendly website: http://sciwiki.lamost.org/Miningdr8/.

Key words: methods: data analysis – surveys – binaries: spectroscopic – stars: carbon – stars: emission-line, Be – novae, cataclysmic variables.

1 INTRODUCTION

Multi-object spectroscopy allows hundreds to thousands of celestial objects to be observed simultaneously. The observation efficiency has been greatly improved and large data sets will be obtained from wide-field spectroscopic surveys. Such large data sets provide us with more samples through which to perform Galactic and extra-galactic research. In addition, it is possible to find unusual, rare or even some unknown types of celestial objects. The Sloan Digital Sky Survey (SDSS: York et al. 2000) is one of the most ambitious and influential surveys in the history of astronomy, which conducts both imaging and spectroscopic surveys over a large area of the sky. Szkody et al. (2002, 2003, 2004, 2005, 2006, 2007, 2009, 2011) have found a total of 297 cataclysmic variable stars (CV stars) from the SDSS Early Data Release (EDR) and Data Releases 1–8 (DR1–DR8). Rebassa-Mansergas et al. (2011) selected white dwarf main-sequence (WDMS) binary candidates by template-fitting all

★E-mail: lal@bao.ac.cn

© 2013 The Authors
Published by Oxford University Press on behalf of the Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024

DR7 spectra, using combined constraints in both χ^2 and signal-to-noise ratio. Li et al. (2012) reported the finding of 13 old metal-poor F-type hypervelocity star candidates from 370 000 selected stars of SDSS DR7. The LAMOST survey (Cui et al. 2012; Zhao et al. 2012; Luo et al. 2012) contains two main parts: the LAMOST ExtraGAlactic Survey (LEGAS) and the LAMOST Experiment for Galactic Understanding and Exploration (LEGUE) survey of Milky Way stellar structure. The unique design of LAMOST enables it to take 4000 spectra in a single exposure to a limiting magnitude as faint as $r = 19$ at resolution $R = 1800$, which is equivalent to the design goal of $r = 20$ for resolution $R = 500$. LAMOST therefore has great potential to survey a large volume of space efficiently for stars and galaxies. Li et al. (2010) searched for metal-poor stars based on low-resolution spectra from the LAMOST data. A very bright ($i = 16.44$) quasar in the redshift desert was discovered by the Guoshoujing Telescope during the LAMOST commissioning observations (Wu et al. 2010). Ren et al. (2013) identified 34 new WDMSs from LAMOST Pilot Survey data. A new double-peaked blazar was found from LAMOST Pilot Survey data by Shi et al. (2013).

In contrast to the simple sample-reduction and template-matching methods, data mining provides a new solution for the discovery of special and unusual types of objects from astronomical survey data sets. Tu et al. (2009, 2010) applied a Local Outlier Factor (LOF) based outlier-detection algorithm to the discovery of supernovae from the SDSS galaxy spectra. The self-organizing maps method was used by Meusinger, Schalldach & Scholz (2011) to select unusual quasars from nearly 10^5 spectra classified as quasars at redshifts from $z = 0.6$–4.3 by the SDSS pipeline. Jiang, Luo & Zhao (2011, 2012) and Jiang et al. (2013) used data-mining methods (including SVM and random forests) to find CV star candidates from the SDSS and LAMOST data sets. Peng et al. (2012) developed a classification system constituted by several Support Vector Machine (SVM) classifiers to select quasar candidates from large sky survey projects to create the input catalogue of quasars for LAMOST or other spectroscopic survey projects.

In this paper, we propose a novel outlier-mining method: using the Monte Carlo Local Outlier Factor (MCLOF) to select unusual and rare spectra. The method efficiently and automatically highlights outliers by marking every spectrum with a number, i.e. the *outlier index*, as a flag for an unusual and rare spectrum. We apply this method to nearly 656 801 stellar spectra (classified as STAR by the SDSS pipeline) from SDSS DR8. In Section 2 we introduce the proposed methods and some other related methods. In Section 3 we show the experiment using SDSS DR8 stellar spectra and the analysis of the results. Finally, a brief conclusion is given in Section 4.

2 METHOD

Our proposed method is based on principal component analysis (PCA), the LOF and the Monte Carlo method. These three methods are first introduced in this section, before defining our proposed method.

2.1 PCA

Principal Component Analysis (PCA: Jolliffe 1986) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (i.e. accounts for as much variability in the data as possible) and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (i.e. uncorrelated with) the preceding components.

Suppose we have a random vector population $\boldsymbol{x}$, where

$$\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}},$$

and the mean value of that population is denoted by

$$\mu_x = E\{\boldsymbol{x}\}$$

and the covariance matrix of the same data set by

$$C_x = E\{(\boldsymbol{x} - \mu_x)(\boldsymbol{x} - \mu_x)^{\mathrm{T}}\}.$$

The components of C_x, denoted by C_{ij}, represent the covariances between the random variable components x_i and x_j. Component C_{ii} is the variance of component x_i. The variance of a component indicates the spread of the component's values around its mean. If two components x_i and x_j of the data are uncorrelated, their covariance is zero ($\mathrm{C}_{ij} = \mathrm{C}_{ji} = 0$). The covariance matrix is, by definition, always symmetric.

From a sample of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$, we can calculate the sample mean and the sample covariance matrix as the estimates of the mean and the covariance matrix. From a symmetric matrix, such as the covariance matrix, we can calculate an orthogonal basis by finding its eigenvalues and eigenvectors. The eigenvectors e_i and the corresponding eigenvalues λ_i are the solutions of the equation

$$C_x e_i = \lambda_i e_i, \qquad i = 1, ..., n.$$

By ordering the eigenvectors in the order of descending eigenvalues (largest first), one can create an ordered orthogonal basis with the first eigenvector having the direction of the largest variance of the data. In this way, we can find the directions in which the data set has the most significant amounts of energy.

Suppose one has a data set, the sample mean and covariance matrix of which have been calculated. Let $\mathbf{A}$ be a matrix consisting of eigenvectors of the covariance matrix as the row vectors. By transforming a data vector $\boldsymbol{x}$, we obtain

$$\boldsymbol{y} = \mathbf{A}(\boldsymbol{x} - \mu_x),$$

which is a point in the orthogonal coordinate system defined by the eigenvectors. Components of $\boldsymbol{y}$ can be seen as the coordinates in the orthogonal base. We can reconstruct the original data vector $\boldsymbol{x}$ from $\boldsymbol{y}$ by

$$\boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{y} + \mu_x,$$

using the property of an orthogonal matrix $\mathbf{A}^{-1} = \mathbf{A}^{\mathrm{T}}$. The original vector $\boldsymbol{x}$ was projected on the coordinate axes defined by the orthogonal basis. The original vector was then reconstructed by a linear combination of the orthogonal basis vectors.

Instead of using all the eigenvectors of the covariance matrix, we may represent the data in terms of only a few basis vectors of the orthogonal basis. If we denote the matrix having the d first eigenvectors as rows by A_d, we can create a similar transformation to that seen above:

$$\boldsymbol{y} = A_d(\boldsymbol{x} - \mu_x)$$

and

$$\boldsymbol{x} = A_d^{\mathrm{T}}\boldsymbol{y} + \mu_x.$$

This means that we project the original data vector on coordinate axes having the dimension d and transform the vector back by

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024

a linear combination of basis vectors. This minimizes the mean-square error between the data and this representation with a given number of eigenvectors.

2.2 Local Outlier Factor

Breunig et al. (2000) presented the conception of the LOF used for outlier detection, which can be applied to describe the singularity of a spectrum in all spectra. The definitions related to the LOF are as follows (Breunig et al. 2000; Tu et al. 2009, 2010).

Definition 1 (*k-distance* of an object p): For any positive integer k, the k-distance of object p, denoted as *k-distance* (p), is defined as the distance $d(p, o)$ between p and an object $o \in D$, such that (i) for at least k objects $o' \in D\backslash p$, it holds that $d(p, o') \leq d(p, o)$ and (ii) for at most $k-1$ objects $o' \in D\backslash p$, it holds that $d(p, o') < d(p, o)$.

Definition 2 (*k-distance* neighbourhood of an object p): Given the *k-distance* of p, the *k-distance* neighbourhood of p contains every object with a distance from p not greater than the *k-distance*, i.e. $N_{k-\text{distance}(p)}(p) = \{q \in D\backslash p | d(p, q) \leq k\text{-distance}(p)\}$. These objects q are called the *k-nearest* neighbours of p. We simplify the notation to use $N_k(p)$ as a shorthand for $N_{k-\text{distance}}(p)$.

Definition 3 (reachability distance of an object p*w.r.t* object o): Let $k \in Z^+$, the reachability distance of object p with respect to object o, be defined as $reachdist_k(p, o) = max\{k\text{-distance}(o), d(p, o)\}$.

Definitions 4 and 5: The local reachability density and the local outlier factor of an object p are defined as

$$lrd_k(p) = \frac{|N_k(p)|}{\sum_{o \in N_k(p)} reachdist_k(p, o)},$$

$$LOF_k(p) = \frac{\sum_{o \in N_k(p)} \frac{lrd_k(o)}{lrd_k(p)}}{|N_k(p)|}.$$

2.3 Monte Carlo methods

Monte Carlo methods are a class of computational algorithms relying on repeated random sampling to compute their results. Application of the Monte Carlo method can usually be roughly divided into two categories: one is to solve the problem with intrinsic randomness so that the computation can simulate this random process, while another is to use random sampling and estimation of the characteristics of random variables as the characteristics of whole samples.

From discussions in Section 2.2, we know that if we want to calculate the LOF of each spectrum in N samples, a N square matrix is needed to store the distances between the samples. It is clear that memory will be the bottleneck when N is large enough. Moreover, it is not necessary and efficient to calculate such a large matrix. Therefore, we use random sampling to obtain a subsample with size n ($n \ll N$) from all samples. We will calculate the LOF in the subsample. After certain iterations, or when the iteration converges, an outlier index will be obtained that can be used to select global outliers from the entire sample.

2.4 The proposed MCLOF method

Combining the above three methods, we define our proposed method as follows.

(i) Randomly select a certain number of spectra with high S/N and use PCA to obtain a feature matrix $\mathbf{T}$ that will be used to perform dimension reduction.

(ii) Project all spectra on to the PCA feature space $\mathbf{T}$ obtained above, then obtain a matrix $\mathbf{D}$.

(iii) Suppose that C_s, C_{LOF} are two vectors with N elements, where N is the count of rows in $\mathbf{D}$.

(iv) Repeat the following iteration a certain number of times:

(a) select a subsample of $\mathbf{D}$ with certain size n ($n \ll N$), denoted by $\mathbf{D}'$;

(b) calculate LOF L_{sub} of $\mathbf{D}'$ and then sort $\mathbf{D}'$ by L_{sub} in ascending order;

(c) add the corresponding sorted index to C_{LOF} and $C_s = C_s + n$.

(v) Now define the outlier index as *outlier index* $= C_{LOF}/C_s$.

(vi) As shown below, spectra with S/N ≥ 3 and *outlier index* ≥ 0.85 are unusual and rare spectra.

(vii) For better classification and analysis of unusual spectra, the clustering method (K-means in our work) is used to divide the selected spectra into different groups.

(viii) The average spectrum of each group is calculated to show obvious spectral features in the group.

3 UNUSUAL STELLAR SPECTRA FROM SDSS DR8

3.1 The data sample

The SDSS DR8 (Aihara et al. 2011), which follows the Early Data Release and Data Releases 1–7 of SDSS-I/II, was made available in 2011 January. The DR8 contains all imaging data taken by the SDSS imaging camera and reproduced by the improved data processing pipeline, as well as new spectra taken by the SDSS spectrograph during its last year of operations for the SEGUE-2 project (Yanny et al. 2011). We select 656 801 stellar spectra (classified as STAR by the SDSS Pipeline) from the SDSS DR8 and about 341 822 objects are from the SEGUE project. By using the spectrum and associated error-estimate vectors (in the form of inverse variances) to derive parameters of interest through χ^2 model fitting to the spectra in pixel space (Glazebrook, Offer & Deeley 1998), each spectrum is labelled with a subclass by the pipeline.

3.2 Results and discussions

We apply the proposed method to the above-mentioned SDSS DR8 stellar spectra data set. Following the steps of the proposed method described in Section 2.4, we will describe and discuss the experimental results in this section.

In order to obtain better eigenspectra, we should select as many spectra with high quality as possible. In our work, a total of 20 000 spectra with S/N > 20 are randomly selected to obtain the eigenspectra using the PCA method. The sum of variance contribution rates of the first 15 eigenspectra (see Fig. 1) is already larger than 99.99 per cent. We can therefore use these 15 eigenvectors to perform data dimension reduction. In order to obtain the *outlier index* for each spectrum, a total of 1 percent of spectra are randomly selected in each iteration and a total of 50 000 iterations carried out. Namely, on average, each spectrum will be selected about 500 times. As expected, all selected times are in the range [420, 580].

After 50 000 iterations, an *outlier index* is calculated for each spectrum. The average of all *outlier index* values is 0.499 and the

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024

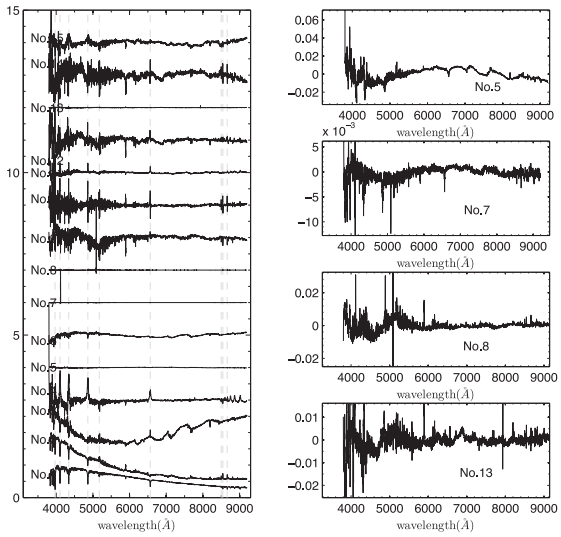

Figure 1. The first 15 eigenspectra obtained from 20 000 randomly selected stellar spectra. The left panel shows all eigenspectra and the fifth, seventh, eighth and thirteenth eigenspectra, which remove the narrow strong lines, are plotted in the right panel.

standard deviation is 0.270. For all spectra with S/N $\geq$ 3, the values are 0.457 and 0.241 respectively. Fig. 2 shows the number density distribution in the grid of S/N and *outlier index*. We choose those 37 033 spectra with S/N $\geq$ 3 and *outlier index* $\geq$ 0.85 as outliers. Fig. 3 shows the spatial distribution of these objects.

The subclass distribution of selected unusual spectra is shown in Table 1. The values $P1 = A1/656\,801$ and $P2 = A3/37\,033$ show the proportion of the corresponding subclass in all samples and mined outliers. For each subclass, $P3 = A3/A2$ represents the outlier trend following our criterion: *outlier index* $\geq$ 0.85. A similar value $P4 = A4/A2$ is calculated to show the outlier trend if we extend the criterion to *outlier index* $>$ 0.7. Despite the fact that the number of objects in each subclass is not very precise due to the misclassification of some spectra, the calculated values are acceptable statistically, especially for spectra with S/N $>$ 3. As excepted, objects with a subclass of Carbon, WD magnetic, L-type, T-type, CV, WD, O-type and B-type have higher $P3$ and $P4$ value than the average value because of the rarity of these types of stars. More than 90 per cent of spectra with a subclass of carbon, WD, L-type, T-type, CV will be selected as outliers if we extend the criterion to *outlier index* $>$ 0.7. Beside these types, some of these outliers are shown to be binary stars, emission-line stars and stars with unusual continuum. The results show that our proposed method can efficiently find these unusual and rare types of spectra from the survey data sets. In addition, some relatively rare and interesting spectra are also selected, indicating that the proposed method can also be used to find rare, even unknown, spectra.

K-means clustering is a method of clustering analysis that aims to divide all samples into k clusters, in which each observation belongs to the cluster with the nearest mean. All the unusual spectra are clustered into 600 groups using the K-means algorithm. A centre spectrum is reconstructed from each clustering group's centre point and an average spectrum is also calculated. The centre spectra and average spectra are intended to show obvious spectral features in each group. We analyse the selected unusual spectra by visual inspection of these spectra group by group. Considering that there are many unusual and rare types of spectra and some rare but interesting single objects that can be picked out from the outliers, a

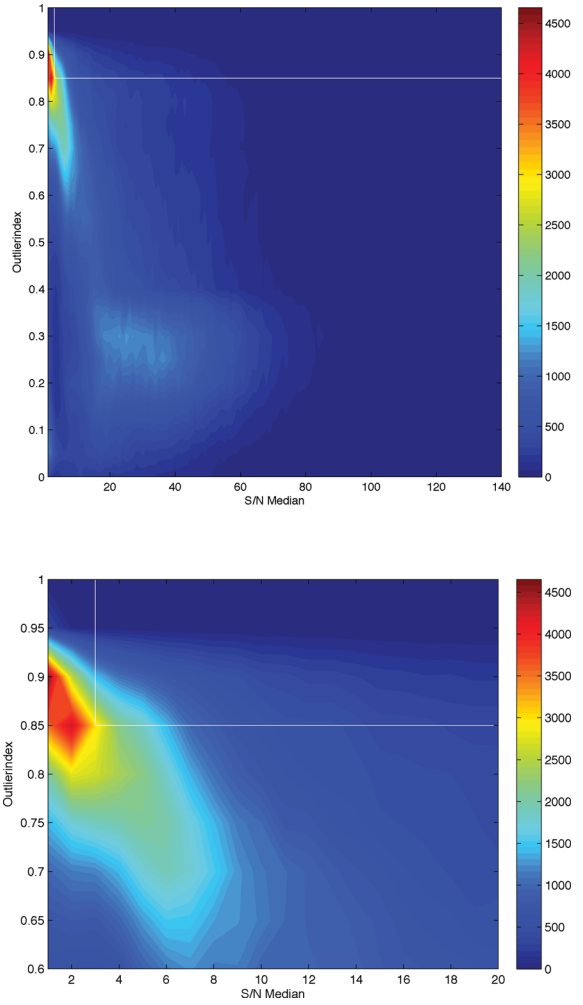

Figure 2. The number density distribution in the grid of S/N and *outlier index* (upper panel). The S/N is from 0−140 in steps of 2. The *outlier index* is from 0–1 in steps of 0.02. The vertical and horizontal white lines are, separately, the lower limits of S/N and *outlier index* for selecting our objects. The area with S/N from 0–20 and *outlier index* from 0.6–1.0 is shown enlarged in the bottom panel.

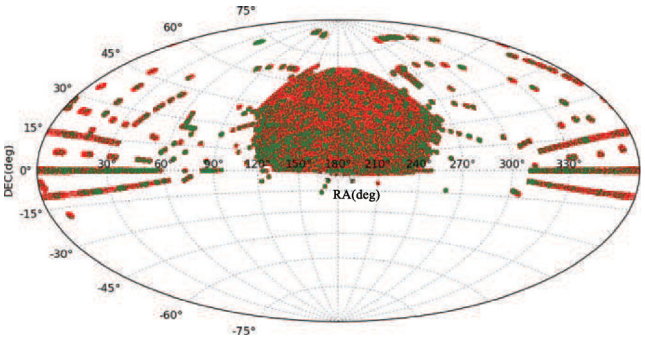

Figure 3. The spatial distribution of our selected abnormal spectra. All (red and green in the online article) dots represent the equatorial coordinates of 656 801 stellar spectra from the SDSS DR8; the green ones in the online article represent our selected abnormal objects.

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024

Table 1. The subclass of all stellar spectra from the SDSS DR8.

Subclass in the SDSS DR8	A1[a]	A2[b]	A3[c]	A4[d]	P1[e]	P2[f]	P3[g]	P4[h]
Carbon, carbon WD, carbon_lines	4731	1170	882	1151	0.72 per cent	2.38 per cent	75.38 per cent	98.37 per cent
WD magnetic	523	148	89	123	0.07 per cent	0.24 per cent	60.14 per cent	83.11 per cent
L1 L2 L3 L4 L5 L5.5 L9	15359	1063	638	977	2.34 per cent	1.72 per cent	60.02 per cent	91.91 per cent
T2	10061	887	504	820	1.53 per cent	1.36 per cent	56.82 per cent	92.44 per cent
CV	4909	2428	819	2190	0.75 per cent	2.21 per cent	33.73 per cent	90.20 per cent
WD	15799	13921	3137	6512	2.41 per cent	8.47 per cent	22.53 per cent	46.78 per cent
O OB B	8247	4143	838	1539	1.26 per cent	2.26 per cent	20.23 per cent	37.15 per cent
M0–M9 M0V M2V	103131	67102	6976	16749	15.70 per cent	18.84 per cent	10.40 per cent	24.96 per cent
A0 A0p	72108	71283	6539	24111	10.98 per cent	17.66 per cent	9.17 per cent	33.82 per cent
K1 K3 K5 K7	120285	114574	6278	24307	18.31 per cent	16.95 per cent	5.48 per cent	21.22 per cent
G0 G2 G5	48211	46956	1889	8065	7.34 per cent	5.10 per cent	4.02 per cent	17.18 per cent
F2 F5 F9	253437	252070	8444	36793	38.59 per cent	22.80 per cent	3.35 per cent	14.60 per cent

[a]Number of stellar spectra among 656 801 DR8 spectra.
[b]Number of stellar spectra with S/N > 3 among 656 801 DR8 spectra.
[c]Number of stellar spectra among 37 033 unusual spectra.
[d]Number of stellar spectra with S/N > 3 and *outlier index* > 0.7 among 656 801 DR8 spectra.
[e]$P1 = A1/656\,801$.
[f]$P2 = A2/37\,033$.
[g]$P3 = A3/A2$.
[h]$P4 = A4/A2$.

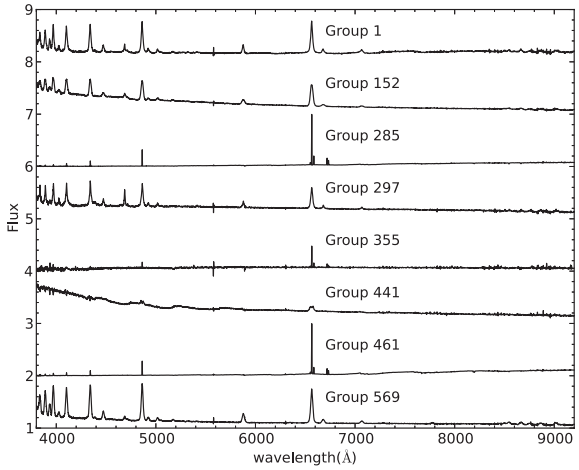

Figure 4. The average spectra of our samples in groups 1, 152, 285, 297, 355, 441, 461 and 569. Each spectrum is linearly normalized into the range [0, 1] when calculating the average spectrum.

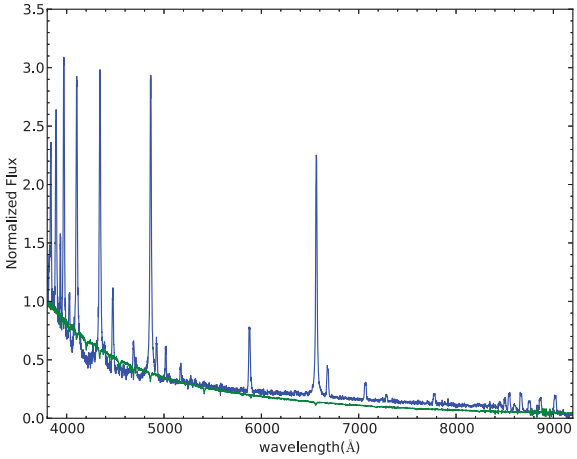

Figure 5. Spectral comparison between a CV star and a normal star. In the online article, the blue line is the spectrum of a CV star and the green line is the spectrum of a normal star.

few typical types and rare objects are chosen for discussion in the following subsections.

3.2.1 Emission-line stars

As is known, the majority of stellar spectra are characterized by obvious absorption lines or absorption bands and only a small number of stellar spectra show obvious emission lines. Stellar spectra with emission lines are usually peculiar objects, such as CV stars, Herbig Ae/Be stars and planetary nebulae. After clustering, some emission-line stars are gathered together into eight groups marked as 1, 152, 285, 297, 355, 441, 461 and 569 separately, the average spectra of which show strong Balmer emission lines (see Fig. 4). Some other emission-line stellar spectra also appear in other groups, the average spectra of which do not show obvious emission lines. We will choose two typical types of spectra with emission lines to discuss in this section.

I: CV stars. Compared with normal stars (see Fig. 5), the spectra of CV stars are these with strong hydrogen Balmer and helium emission lines that typically signify ongoing accretion. Szkody et al. (2002, 2003, 2004, 2005, 2006, 2007, 2009, 2011) have found a total of 297 CV stars from the SDSS EDR, DR1–DR8, of which 271 objects have spectra. Among these 271 objects, there are 248 spectra classified as STAR by the SDSS pipeline and with S/N > 3. After cross-matching with Szkody's catalogue, we find that there are 226 CV stars in our selected outliers, while only 22 (∼8 per cent) do not appear in our selected outliers. By checking these spectra, we find that the outlier indexes of 10 spectra are larger than 0.7; also the spectra of 12 other objects show strong noise around the weak emission lines. After visual inspection of the groups that have spectra of CV stars, we find 33 new CV stars not included in Szkody's catalogue; 26 objects are the sources in the catalogue observed twice or more times (see Table 2). SDSS J143947.62−010606.8 (observed twice) is identified as a CV star by Shimansky, Bikmaev & Shimanskaya (2011). The remaining five sources SDSS J035747.16−063850.7, SDSS J140429.37+172359.4, SDSS J143209.78+191403.5, SDSS J111715.90+175741.7 and SDSS J161935.76+524631.8 are our

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024

Table 2. The list of CV stars observed more than twice.

Designation	MJD	Plate	Fiber id
SDSS J143500.22−004606.3	51637	306	4
	51690	306	20
SDSS J143947.62−010606.9	51663	307	90
	52409	919	620
SDSS J163722.21−001957.1	51671	348	103
	51696	348	103
SDSS J173008.39+624754.6	51694	352	26
	51789	352	33
SDSS J172601.95+543230.7	51813	357	51
	51997	367	53
SDSS J223439.92+004127.2	52143	376	631
	52201	674	427
SDSS J223843.83+010820.6	52145	377	540
	52201	674	524
SDSS J085344.17+574840.6	51902	483	332
	51924	483	348
	51942	483	352
SDSS J132723.38+652854.2	51973	496	83
	51988	496	83
SDSS J204448.91−045928.8	52145	635	387
	53269	1916	424
SDSS J013132.38−090122.2	52178	662	384
	54465	2878	585
	52147	662	384
SDSS J073817.74+285519.6	52232	754	34
	52339	888	378
SDSS J094325.89+520128.7	52281	768	13
	53763	2384	7
SDSS J161030.34+445901.7	52355	814	274
	52370	814	274
	52443	814	280
SDSS J074813.54+290509.1	52592	1059	405
	52618	1059	407
SDSS J092444.48+080150.9	52724	1195	403
	54176	2402	409
SDSS J155412.33+272152.4	53498	1654	542
	54544	2459	521
	54339	2459	533
SDSS J155904.62+035623.4	53494	1837	283
	54592	2951	65
SDSS J090950.53+184947.4	53687	2285	30
	53700	2285	30
SDSS J032855.00+052254.1	53713	2334	581
	53730	2334	600
SDSS J105754.25+275947.5	53800	2359	102
	53826	2359	105
SDSS J112003.40+663632.4	54464	2858	228
	54498	2858	239

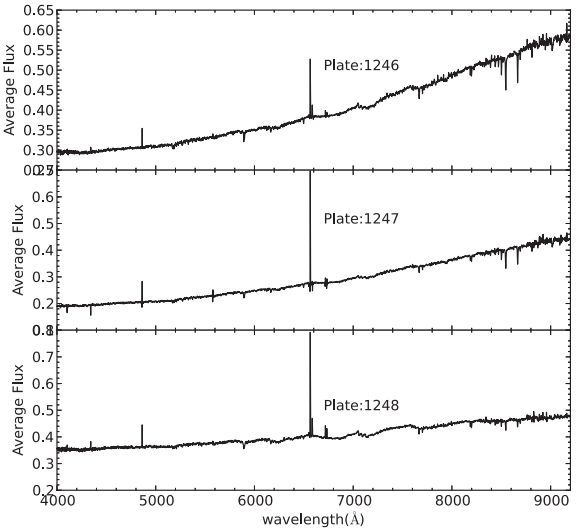

Figure 6. The average spectra of our samples in plates 1246, 1247 and 1248. Each spectrum is linearly normalized into the range [0, 1] when calculating the average spectrum.

first discovered CV stars. Their basic information is shown in Table 3.

II: nebulae. Since a nebula is an interstellar cloud of dust, hydrogen, helium and other ionized gases, the spectra show very strong H and He emission lines. Stars located behind the nebula in sight will also show emission lines in their spectra. We note that the selected spectra in plates 1246, 1247 and 1248 show strong Balmer emission lines (see Fig. 6). From the imaging of this zone (see Fig. 7), the nebula is the major cause of the emission lines in the spectra.

Due to the strong emission lines, the two types of stars are selected as outliers from normal stars. After visual inspection of all clustering groups, we find that there are also some other types of emission-line spectra. These results indicate that our method is very useful in quickly identifying stellar spectra with emission lines.

3.2.2 Double stars

A binary star is a star system consisting of two stars orbiting around their common centre of mass (Bagnuolo & Gies 1992; Batten 1973).

Here, we take the white dwarf main-sequence (WDMS) binary as an example to show our results for double stars. Binary stars containing a white dwarf primary and a main-sequence companion (WDMS) were initially main-sequence binaries in which the more massive star evolved through the giant phase and became a white dwarf (Rebassa-Mansergas et al. 2011). Rebassa-Mansergas et al. (2011) selected WDMS binary candidates by template-fitting all 1.27 million SDSS DR6 spectra, using combined constraints in both

Table 3. Basic information for the CV stars first uncovered in this paper from the SDSS DR8.

Designation	MJD	Plate	Fiber id	g^a (mag)	$u-g^b$	$g-r$	$r-i$	$i-z$
SDSS J035747.16−063850.7	53741	2071	137	19.99664	−0.12799	0.2604	−0.13382	0.14473
SDSS J140429.37+172359.4	54509	2757	112	17.54387	−0.42746	0.18133	−0.01964	0.1285
SDSS J143209.78+191403.5	54534	2774	107	18.39235	0.14915	0.04772	−0.01246	−0.04828
SDSS J111715.90+175741.7	54951	3327	337	–[c]	–	–	–	–
SDSS J161935.76+524631.8	54983	3442	276	19.01755	−0.15327	0.16525	0.35507	0.21506

[a] g represents the g-band magnitude given by the SDSS DR8.
[b] $u-g, g-r, r-i$ and $i-z$ respectively represent four different colours.
[c] – represents no photometric data given by the SDSS DR8.

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024

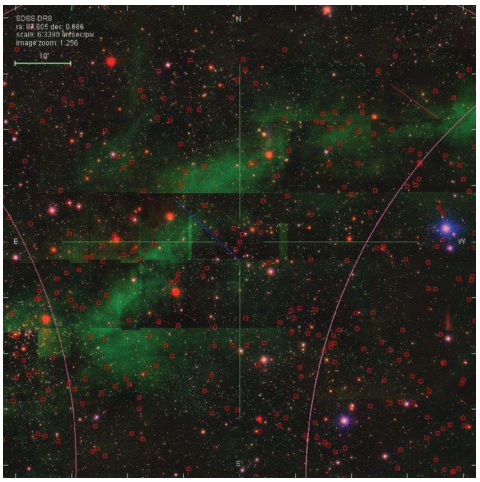

Figure 7. The SDSS image of the region of plates 1246, 1247 and 1248. The centre of this area is RA = 87.805°, Dec. = 0.686°. The red squares in the online article denote objects with emission lines.

χ^2 and S/N. They found 2248 candidates from about 70 000 spectra by visual inspection. Among those spectra, we find that 2091 spectra with S/N $\geq$ 3 are classified as 'Star' by the SDSS DR8 pipeline and 2074 are selected as outliers with our proposed MCLOF method. In other words, 99.18 per cent of these objects can be picked out successfully. We find that most spectra not mined by our proposed method are missed because one of the two components in the spectra is much weaker than the other. Because of their high similarity, these spectra are gathered in groups after clustering. There are 50 groups in total, containing 1914 WDMS spectra. We manually check the group containing WDMS objects and easily find ∼500 more new WDMS candidates not included in Rebassa-Mansergas et al. (2011). Obviously, our proposed MCLOF method is much more efficient in finding WDMS binaries than that of Rebassa-Mansergas et al. (2011).

There are also some spectra consisting of two different physically irrelevant components (usually called an optical pair), which are not bound and are simply along the same line of sight. Due to the fibre radius and observation conditions (i.e. seeing, wind and others), the light from two or even more objects is collected into the same fibre. In our selected objects, there are some spectra of this type. Fig. 8 shows a spectrum consisting of a QSO and an M-type star. This is obviously not a physical binary.

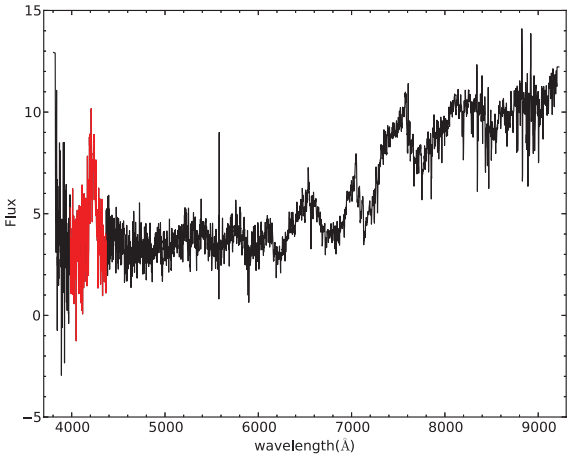

Figure 8. The spectrum of SDSS J031315.40+001036.4. The black line is its spectrum and the shaded one (red in the online article) the C IV broad line in the QSO spectrum.

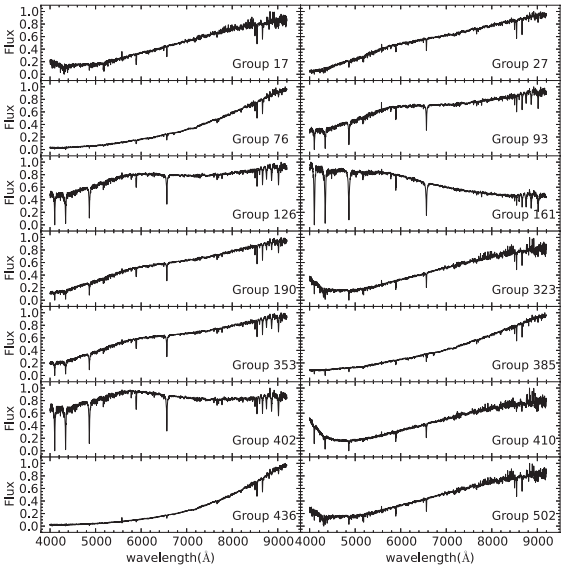

Figure 9. The average spectra of our sample in groups 17, 27, 76, 93, 126, 161, 190, 323, 353, 385, 402, 410, 436 and 502. Each spectrum is linearly normalized into the range [0, 1] when calculating the average spectrum.

3.2.3 *Stellar spectra with an unusual continuum*

There are 14 clustering groups (Fig. 9 shows their average spectra) with strongly unusual continua. We obviously find that their continua are very different from those in normal spectra with the same spectral lines. We find that spectra with unusual continua always appear in the same plate and these spectra are also clustered in the same groups. Compared with normal stellar spectra, the continua of some spectra are unusual because of interstellar extinction or bad flux calibration.

We plot randomly selected outliers in plates 1246, 1247, 1638, 1640, 2555 and 2887 in Fig. 10. We find that the physical flux

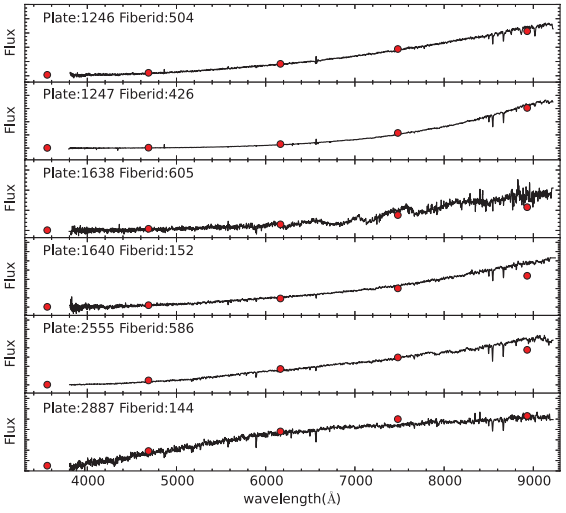

Figure 10. Randomly selected spectra in plates 1246, 1247, 1638, 1640, 2555 and 2887. The black solid lines are their spectra and the filled circles (red in the online article) are physical fluxes converted from the SDSS *ugriz* magnitudes.

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024

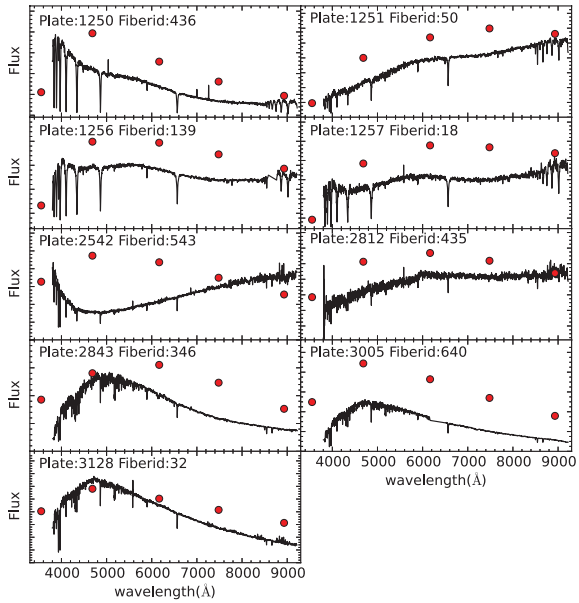

Figure 11. Randomly selected spectra in plates 1250, 1251, 1256, 1257, 2542, 2812, 2843, 3005 and 3128. The black solid lines are their spectra and the filled circles (red in the online article) are physical fluxes converted from the SDSS *ugriz* magnitudes.

converted from the magnitudes of these objects is consistent with the spectral flux. These spectra with unusual continuum are therefore caused by high interstellar extinction.

The randomly selected outliers in plates 1250, 1251, 1256, 1257, 2542, 2812, 2843, 3005 and 3128 are plotted in Fig. 11. We find that the physical flux converted from their magnitudes is not in agreement with the spectral flux, which should be exactly equal to the physical flux. Therefore, we conjecture that the unusual continua of these spectra has possibly been changed by bad flux calibration.

3.2.4 Carbon stars, L-type stars and T-type stars

Carbon stars are late-type stars similar to red giant stars (or occasionally to red dwarfs), the atmospheres of which contain more carbon than oxygen. Their spectra (see Fig. 12) are dominated by strong molecular bands due to molecules including carbon, namely CH, CN and C_2.

The SDSS pipeline classified some spectra as 'Carbon Star', 'Carbon WD' or 'Carbon_lines'. As shown in Table 1, 75.38 per cent of these objects are selected as outliers. Approximately ∼98.37 per cent of these objects will be selected as outliers if we extend the criterion to *outlier index* > 0.7. Similarly to WDMS, carbon stars are gathered in groups due to their high similarity and there are eight groups containing more than 60 spectra of this type among the clustering groups. All the average spectra (see Fig. 13) of these groups show obvious CN and C features.

Similarly, the L-type and T-type stars are those with lower luminosity and effective temperature than M9.5-type stars. We find that more than 92.3 per cent of objects of these two types have low quality (S/N < 3). About 58.5 per cent of those objects with S/N ≥ 3 are picked out as outliers. About 92.1 per cent of these objects will be selected as outliers if we extend the criterion to *outlier index* > 0.7. After visual inspection of the spectra, we find that most of the spectra with S/N ≥ 3 are misclassified and the real T-type and L-type stellar spectra have very low S/N.

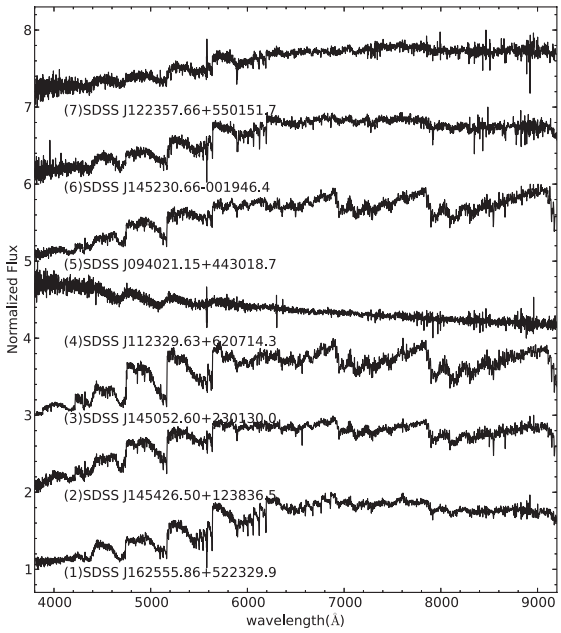

Figure 12. The spectra of seven carbon stars selected as outliers.

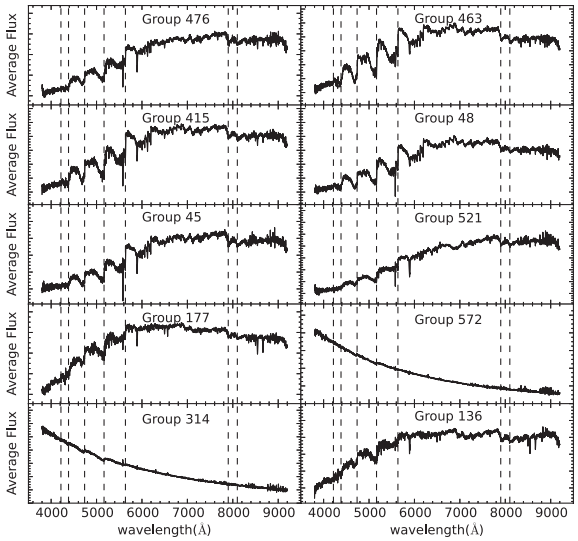

Figure 13. The average spectra for groups 476, 463, 415, 48, 45, 521, 177, 572, 314 and 136 of our sample. Each spectrum is linearly normalized into the range [0, 1] when calculating the average spectrum.

3.2.5 White dwarfs (WD)

White dwarf stars (WD stars), also called degenerate dwarfs, are stellar remnants composed mostly of electron-degenerate matter (Eisenstein et al. 2006). Their spectra are dominated by very broad H Balmer lines or He lines. As shown in Table 1, 14 609 spectra are classified as WD stars by the SDSS pipeline. Although WD stars are usually considered as rare objects, the total number of WD stars in the SDSS is not small due to target-selection effects and only 3226 (∼22.5 per cent) of objects are selected as outliers. Similarly, only 2209 objects (∼25.2 per cent) from 8758 spectra with S/N ≥ 3 are picked out as outliers after cross-matching our selected unusual spectra with 10 555 spectroscopically confirmed white dwarfs from the SDSS DR4 in Eisenstein et al. (2006). However, the proportion for particular types of WD stars is much larger than total proportion.

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024

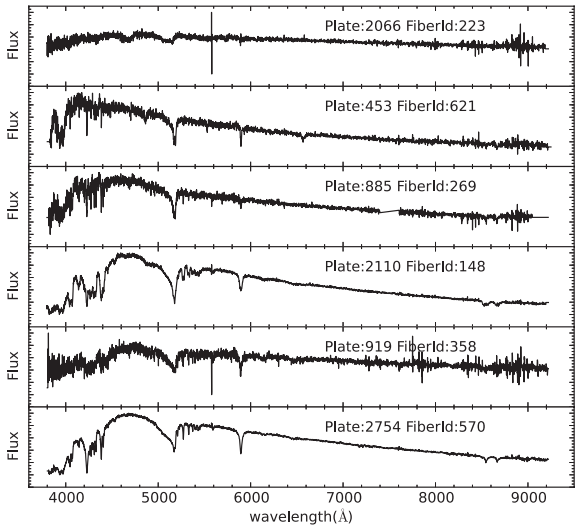

Figure 14. The spectra of six DZ white dwarf stars that were listed in Koester et al. (2011) and are also in our outliers.

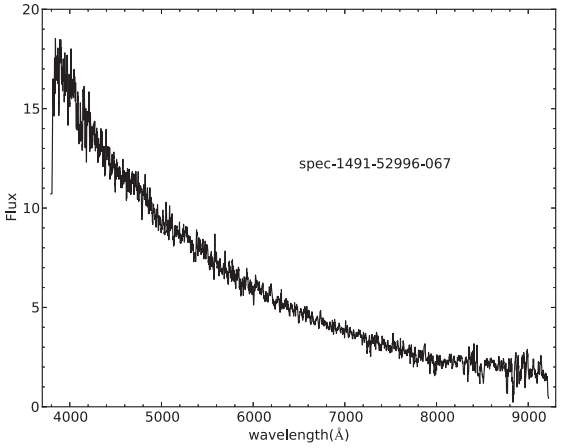

Figure 15. The spectra of a DC white dwarf star.

We will discuss three special types (DZ, DC and DQ) of WD stars here.

I: DZ-type WD stars. Cool white dwarfs with traces of metals other than carbon are designated with the letter *Z* in the current classification system (Eisenstein et al. 2006). There are only metal lines and no H or He lines in these types of spectra. There are 25 spectra labelled DZ with S/N $\geq$ 3 in Eisenstein et al. (2006)'s catalogue and 21 (84 per cent) spectra are selected as outliers. Koester et al. (2011) found 26 spectra, of which six spectra are classified as 'Star' by the SDSS Pipeline. Those six spectra (see Fig. 14) are all in our outliers and the *outlier index* of five spectra is larger than 0.92.

II: DC-type WD stars. A DC white dwarf is another type of cool white dwarf star. A spectrum of this type is shown in Fig. 15 as only a continuous spectrum; there are no lines deeper than 5 per cent in any part of the electromagnetic spectrum (McCook & Sion 1999). Among the objects in Eisenstein et al. (2006), there are 332 DC white dwarf stars and 121 objects with S/N $\geq$ 3 are classified as 'Star' by the SDSS pipeline. Of these 121 objects, 91 ($\sim$75 per cent) are selected as outliers by our method. We also find that about 114 objects ($\sim$94.2 per cent) will be selected as outliers if the lower limit of *outlier index* is extended to 0.75. The remaining seven objects not selected by our method are plotted in Fig. 16. We can see that three

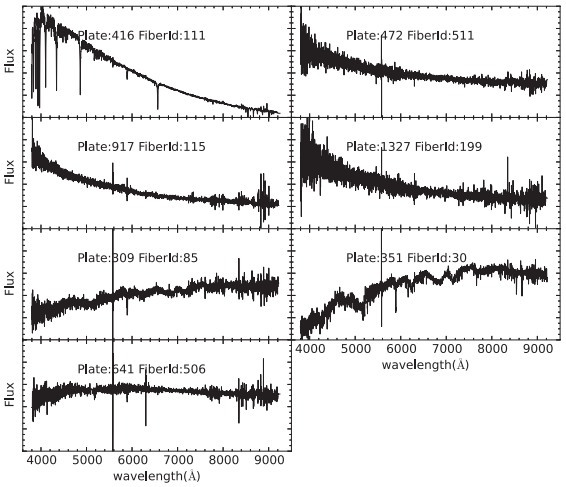

Figure 16. The spectra of seven DC white dwarf stars that are listed in Eisenstein et al. (2006) but do not appear in our unusual spectra with *outlier index* $<$ 0.75.

spectra are obviously not DC white dwarf stars and the remaining four spectra are not selected simply because of their low S/N.

III: DQ-type WD stars. DQ white dwarfs show atomic or molecular carbon features in their spectra. The atmosphere consists of helium and the carbon is believed to be dredged up to the surface by the deepening helium convection zone (Koester & Knist 2006). There are 49 spectra labelled DQ with S/N $\geq$ 3 in Eisenstein et al. (2006)'s catalogue and 44 spectra ($\sim$89.7 per cent) are selected as outliers by our proposed method.

From the above, we can conclude that our method can efficiently select these three and other relatively rare types of white dwarfs.

3.2.6 *Some rare objects*

After visual inspection, we find some rare and interesting spectra in the outliers. There are so many rare and unusual spectra that we cannot explain each spectrum in detail. Some typical objects are discussed in detail in this subsection and more rare spectra will be selected from our outliers in future.

I: SDSS J100811.87+162450.4. The spectrum of this object is shown in Fig. 17. The spectrum contains two components: a white

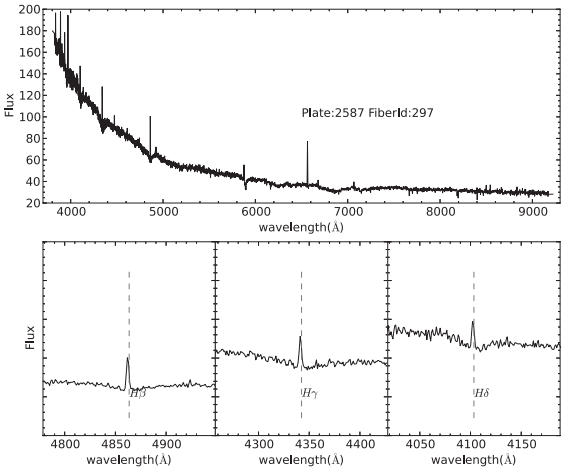

Figure 17. The spectrum of SDSS J100811.87+162450.4 (upper panel). The regions of Hβ, Hγ and Hδ lines in the rest frame are depicted in the bottom panel.

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024

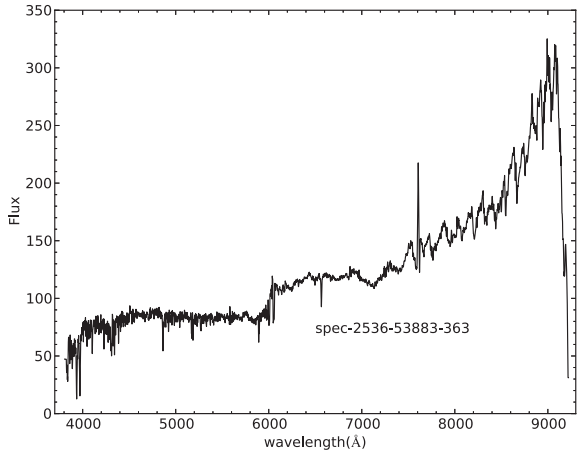

Figure 18. The spectrum of SDSS J190436.74+401022.8.

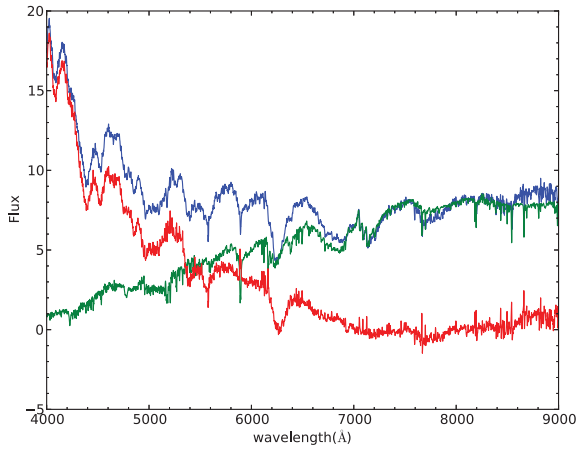

Figure 19. The spectrum of SDSS J120403.01+201443.7. In the online article the blue line is its spectrum, the green line the spectrum of an M2-type model spectrum and the red line the residual spectrum obtained by subtracting the M2 spectrum from the original spectrum.

dwarf and a M-type star. In addition, there are obvious emission lines with different radial velocity in the absorption core.

II: SDSS J190436.74+401022.8. The spectrum of this object is shown in Fig. 18. The spectrum is very unusual and no literature has discussed this object.

III: SDSS J120403.01+201443.7. We also find SDSS J120403.01+201443.7 a very interesting object and its spectrum is shown in Fig. 19. We could find a supernova component ($z = 0.02445$) after subtracting a M2 model spectrum. The SDSS image (see Fig. 20, left panel) is taken before explosion and close to the nearest M2 star (RA = 181.°0125594, Dec. = 20.°24547228) but the spectrum is at the SN 2008bf location RA = 181.°01262, Dec. = 20.°24545 and is a typical SNIa spectrum. The *Hubble Space Telescope* (*HST*) image (see Fig. 20, right panel) is taken after explosion. The centre of the image is obviously brighter than before the explosion.

IV. SDSS J093100.51+485935.7. The spectrum of this object is shown in Fig. 21. A starburst galaxy spectrum ($z = 0.02$) is shown after subtracting a M2-spectype model. From the SDSS image (see Fig. 22), we can obviously see a galaxy behind the star. There are also two similar objects: SDSS J132151.51+420014.1 and SDSS J085125.64+024958.9.

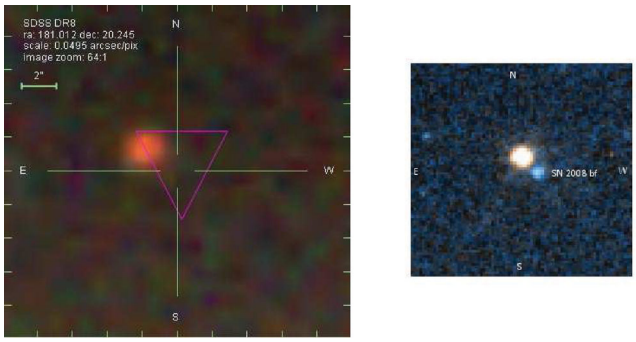

Figure 20. The image of SDSS J120403.01+201443.7. The left panel is the SDSS image taken before the explosion and the right panel is the *HST* image taken after the explosion.

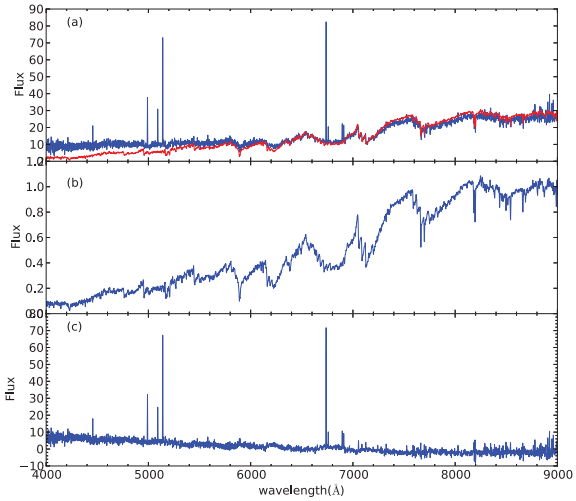

Figure 21. The spectrum of SDSS J093100.51+485935.7. In panel (a), the black line is its spectrum and the shaded line (red in the online article) is a M2-type model spectrum. The model spectrum is replotted in panel (b). Panel (c) is the residual spectrum obtained by subtracting the M2 spectrum from the original spectrum.

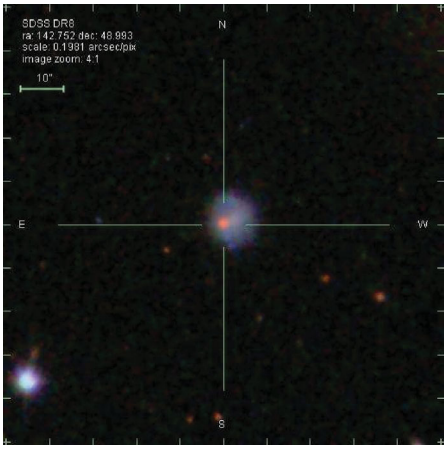

Figure 22. SDSS image of SDSS J093100.51+485935.7. Its centre is RA = 142.°752, Dec. = 48.°993.

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024

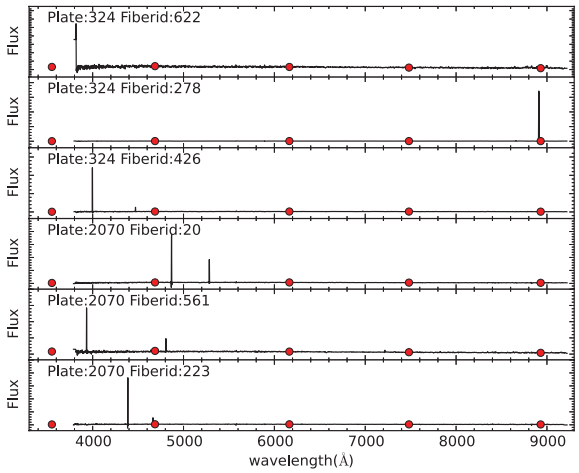

Figure 23. Six selected spectra in plates 324 (MJD: 51616) and 2070 (MJD: 53405). The black solid lines are their spectra and the filled circles (red in the online article) are physical fluxes that are converted from the SDSS *ugriz* magnitudes.

3.2.7 Other objects

There are also some other types of spectra in our outliers. Due to target-selection effects, the total number of F-type, G-type, K-type and M-type stars is much larger than that of other types. Consequently, it is also normal that those types of stars constitute the majority of the selected outliers. In spite of these facts, the selection portion for these types of stars is obviously much smaller than for other types, as shown in Table 1. However, the selected outliers of these types are also abnormal compared with the majority of the corresponding type of stars.

In addition, we find that two plates 324 (MJD: 51616) and 2070 (MJD: 53405) have many outliers. Six spectra selected from these two plates are plotted in Fig. 23. We find that each spectrum has a strong emission line, possibly caused by bad observation conditions or data-reduction errors.

Considering that some spectra of galaxies or QSOs may be classified as stars by the SDSS pipeline, these two types of spectra are also selected as outliers. We find that these spectra are misclassified mainly due to their low S/N.

3.3 A web-based catalogue

In order to publish our results and view the results conveniently, we store the catalogue in the data base. Based on the data base, some web pages (http://sciwiki.lamost.org/Miningdr8/) give web interfaces to access these outliers. There are three ways to access our results: they can be browsed by groups, by plates and by subclasses. The three ways are very similar and we will take the first as an example in order to introduce our web pages. The page viewed by group (see Fig. 24) gives a list of all clustering groups and the average spectrum of each group. One can access the details page (see Fig. 25) of a group by clicking the link text or the average spectrum image. All the unusual spectra in the group are listed in descending order of *outlier index* and the average spectra and object distribution figure are also given. The basic information for each spectrum will be shown when the mouse is moved over the spectrum image. By clicking the link text or the spectrum image, one can access the page (see Fig. 26) of an object for detailed analysis. In this web page we can scan the spectrum, the SDSS imaging and the explorer page of

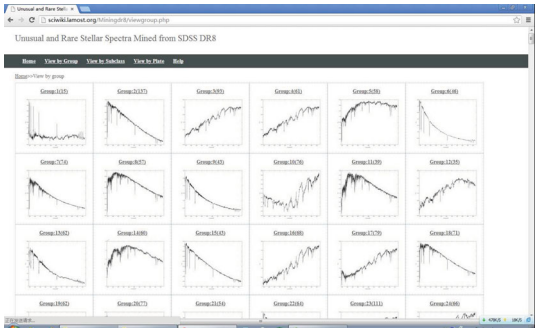

Figure 24. The web page browsed by groups.

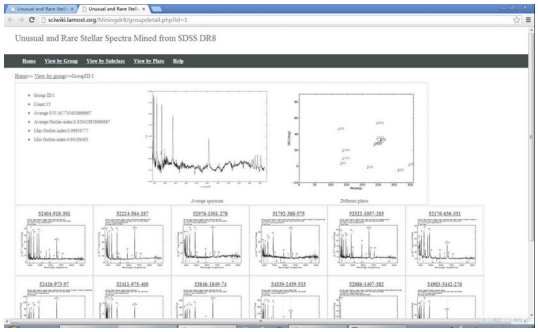

Figure 25. The web page of each clustering group.

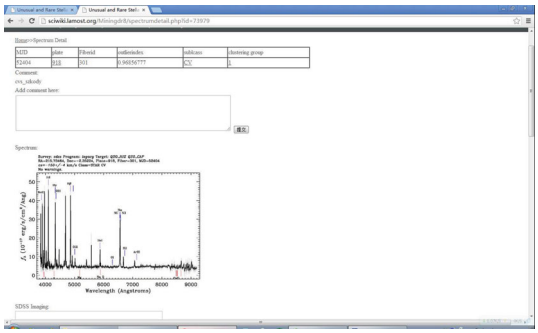

Figure 26. The web page of each spectrum.

the SDSS Catalog Archive Server (CAS) of an object and also view and make comments on the spectrum. Similarly, one can click the link text to view each plate or subclass in detail and the web page of each plate or subclass is similar to that of each group.

4 CONCLUSIONS

In this paper, we propose a novel outlier-detection method, MCLOF, to find unusual and rare spectra from large spectral survey data sets and apply this method to all stellar spectra from the SDSS DR8. We discuss the selected outliers in detail and the results show that our proposed method can quickly and efficiently select unusual spectra such as emission-line stellar spectra, double stars, carbon stars and other types of stellar spectra. Compared with other works based on simple sample reduction and following visual inspection, our method is able to select not only more than 95 per cent of some known special types of spectra such as CV stars and WDMS but also some kinds of spectra left out in other works. In addition, some rare spectra are selected as well, indicating that our method can find rare, even unknown, spectra in large survey data sets. The proposed

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024

method can also be used in other survey data sets and is not restricted only to spectral survey. We provide details of all outliers on the web page http://sciwiki.lamost.org/Miningdr8/.

ACKNOWLEDGEMENTS

The authors thank the anonymous referees for their constructive comments and are grateful to Professors Hu Jingyao, George Comte, Liu Jifeng and Yang Haifeng, Drs Du Wei and Song Yihan, Mr Si Jianmin, Ms Ren Juanjuan and Ms Hou Wen for useful suggestions regarding the methods and analysis of the results. This work is supported by the National Natural Science Foundation of China (Grant Nos 10973021, 61202315, 11078013 and 11203045).

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation and the US Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration, including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, University of Cambridge, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington and Yale University.

REFERENCES

Aihara H. et al., 2011, ApJS, 193, 29
Bagnuolo W. G., Gies D. R., 1992, in McAlister H. A., Hartkopf W. I., eds, ASP Conf. Ser. Vol. 32, Complementary Approaches to Double and Multiple Star Research. Astron. Soc. Pac., San Francisco, p. 140
Batten A. H., 1973, Binary and Multiple Systems of Stars. Pergamon Press, Oxford
Breunig M. M., Kriegel H.-P., Ng R. T., Sander J., 2000, in Chen W., Naughton J. F., Bernstein P. A., eds, Proc. 29th ACM SIDMOD International Conference on Management of Data (SIGMOD 2000). ACM, New York, p. 427
Cui X. et al., 2012, Res. Astron. Astrophys., 12, 1197
Eisenstein D. J. et al., 2006, ApJS, 167, 40
Glazebrook K., Offer A. R., Deeley K., 1998, ApJ, 492, 98
Harris H. C. et al., 1998, ApJ, 502, 437
Jiang B., Luo A. L., Zhao Y. H., 2011, Spectroscopy and Spectral Anslysis, 31, 2278
Jiang B., Luo A. L., Zhao Y. H., 2012, Spectroscopy and Spectral Anslysis, 32, 510
Jiang B., Luo A. L., Zhao Y. H., Wei P., 2013, MNRAS, 429, 4
Jolliffe I. T., 1986, in Jolliffe I. T., ed., Principal component analysis. Springer, Berlin
Koester D., Knist S., 2006, A&A, 454, 951
Koester D., Girven J., Gänsicke B. T., Dufour P., 2011, A&A, 530, 11
Li H. N., Zhao G., Christlieb N. et al., 2010, Res. Astron. Astrophys., 10, 753
Li Y., Luo A. L., Zhao G., Lu Y. J., Ren J. J., Zuo F., 2012, ApJ, 744, L24
Luo A. et al., 2012, Res. Astron. Astrophys., 12, 1243
McCook G. P., Sion E. M., 1999, ApJS, 121, 1
Meusinger H., Schalldach P., Scholz R.-D., 2011, A&A, 541, 27
Peng N., Zhang Y. X., Zhao Y. H., Wu X. B., 2012, MNRAS, 425, 2599
Rebassa-Mansergas A., Nebot Gómez-Morán A., Schreiber M. R., Girven J., Gänsicke B. T., 2011, MNRAS, 402, 620
Ren J. et al., 2013, AJ, submitted
Shi Z. X. et al., 2012, A&A, submitted
Shimansky V. V., Bikmaev I. F., Shimanskaya N. N., 2011, Astrophys. Bull., 4, 442
Szkody P. et al., 2002, ApJ, 123, 430
Szkody P. et al., 2003, ApJ, 126, 1499
Szkody P. et al., 2004, ApJ, 128, 1882
Szkody P. et al., 2005, ApJ, 129, 2386
Szkody P. et al., 2006, ApJ, 131, 973
Szkody P. et al., 2007, ApJ, 134, 185
Szkody P. et al., 2009, ApJ, 137, 4011
Szkody P. et al., 2011, ApJ, 142, 181
Tu L. P., Luo A. L., Wu F. C., Wu C., Zhao Y. H., 2009, Res. Astron. Astrophys., 9, 635
Tu L. P., Luo A. L., Wu F. C., Zhao Y. H., 2010, Science China Physics, Mechanics & Astronomy, 53, 1928
Wu X. B. et al., 2010, Res. Astron. Astrophys., 10, 745
Yanny B. et al., 2011, AJ, 137, 4377
York D. G. et al., 2000, AJ, 120, 1579
Zhao G., Zhao Y. H., Chu Y. Q., Jing Y. P., Deng L. C., 2012, Res. Astron. Astrophys., 12, 723

APPENDIX A: THE UNUSUAL SPECTRA TABLE DESCRIPTION IN THE CATALOGUE DATA BASE

Table A1. Table description for selected spectra in the catalogue data base.

Field[a]	Type[b]	Field description
autoid	int	
mjd	int	SDSS MJD
plate	int	SDSS Plate
fiberid	int	SDSS Fiberid
specobjid	varchar	SDSS Specobjid
raobj	double	ra of this object
decobj	double	dec of this object
subclass	char (20)	subclass given by SDSS CAS
objType	char (20)	object target type
snMedian	double	sn median of the spectra
psfMag_u	double	psfMag_*u* of the object
psfMag_g	double	psfMag_*g* of the object
psfMag_r	double	psfMag_*r* of the object
psfMag_i	double	psfMag_*i* of the object
psfMag_z	double	psfMag_*u* of the object
latobj	double	Galactic latitude of the object
lonobj	double	Galactic longitude of the object
outlierindex	double	Calculated *outlier index*
clustergroup	int	The clustering group id
comment	varcha (250)	Manual comment about the object

[a]Column name in the data base.
[b]The type of value stored in the data base.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/431/2/1800/1466626 by guest on 16 February 2024