SCIENCE CHINA
Physics, Mechanics & Astronomy

• Article •

June 2013 Vol.56 No.6: 1227–1234
doi: 10.1007/s11433-013-5083-8

A SVM-kNN method for quasar-star classification

PENG NanBo[1,2*], ZHANG YanXia[1*] & ZHAO YongHeng[1]

[1] *Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100049, China;*
[2] *Graduate University of Chinese Academy of Sciences, Beijing 100049, China*

Received March 28, 2012; accepted April 11, 2012; published online May 3, 2013

We integrate *k*-Nearest Neighbors (kNN) into Support Vector Machine (SVM) and create a new method called SVM-kNN. SVM-kNN strengthens the generalization ability of SVM and apply kNN to correct some forecast errors of SVM and improve the forecast accuracy. In addition, it can give the prediction probability of any quasar candidate through counting the nearest neighbors of that candidate which is produced by kNN. Applying photometric data of stars and quasars with spectral classification from SDSS DR7 and considering limiting magnitude error is less than 0.1, SVM-kNN and SVM reach much higher performance that all the classification metrics of quasar selection are above 97.0%. Apparently, the performance of SVM-kNN has slighter improvement than that of SVM. Therefore SVM-kNN is such a competitive and promising approach that can be used to construct the targeting catalogue of quasar candidates for large sky surveys.

classification, stars/quasars, algorithm:SVM, kNN, data analysis

PACS number(s): 98.52.Cf, 97.20.-w, 98.54.-h, 07.05.Kf

Citation: Peng N B, Zhang Y X, Zhao Y H. A SVM-kNN method for quasar-star classification. Sci China-Phys Mech Astron, 2013, 56: 1227–1234, doi: 10.1007/s11433-013-5083-8

1 Introduction

Over the years, the volume of astronomical data at different wavebands has grown dramatically with large space-based and ground-based telescopes surveying the sky, such as SDSS, 2MASS, NVSS, FIRST and 2dF. To preselect scientific targets from the enormous amount of observed data is a significant and challenging issue. Thus how to extract knowledge from a huge volume of data by automated methods is a critical task for astronomers. In the next decade, the ongoing or planned multiband photometric survey projects, for instance, the Large Synoptic Survey Telescope (LSST) [1], the Visible and Infrared Survey Telescope for Astronomy (VISTA) [2], and the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) [3] will bring more serious challenges.

Many data mining algorithms have been taken to find quasars in astronomy because traditional quasar selection relies on cut-off in two-dimensional color space despite the fact that most modern surveys are done in several bandpasses. Abraham et al. [4] used a Difference Boosting Neural Network (DBNN) classifier which is a bayesian supervised learning algorithm to formulate a catalogue of quasar candidates. Carballo et al. [5] obtained a sample set of redshift $z \geqslant 3.6$ radio quasi-stellar objects using Neural Network (NN). For these methods based on NN, they usually obtain a perfect performance on artificial test sample and undergo the risk of overfitting. Zhang et al. [6] demonstrated that Support Vector Machine (SVM) showed better performance than learning vector quantization (LVQ) and single-layer perception (SLP) when preselecting AGN candidates. The advantage of SVM is robustness and generalization and it can overcome the weakness of NN methods. The problem, however, of SVM is that it cannot avoid the misclassifica-

*Corresponding author (PENG NanBo, email: nbpeng@bao.ac.cn; ZHANG YanXia, email: zyx@bao.ac.cn)

tion near the hyperplane of SVM. The most representative work, however, could be the series of work completed by Richards et al. [7–10] using bayesian selection based on Kernel Density Estimation (KDE). The clear advantage of this method is that it can effectively combine data mining technology and physical principles such as prior knowledge of celestial objects or their distributions. When this method deals with high-dimensional problems or the prior knowledge of astronomical problem are not ideal, it cannot effectively solve them.

SVM provides a good out-of-sample generalization and can be robust, even when the training sample has some bias. This distinguishing feature of SVM attracts many astronomers to use it on selecting quasar candidates. Gao et al. [11] compared the performance of SVM with *K*-Dimensional Tree (KD-Tree) to separate quasars from stars. Bailer-Jones et al. [12] developed and demonstrated a probabilistic method for classifying quasars in surveys, using the Discrete Source Classifier (DSC), a supervised classifier based on SVM. Kim et al. [13] presented a new QSO selection algorithm using SVM, on a set of extracted times series features including period, amplitude, color and autocorrelation value. Although SVM can effectively discriminate most of unknown objects, if unresolved objects nearly exist in the hyperplane of SVM, the result will become unreliable. Actually, we can add some other methods such as *k*-Nearest Neighbors (kNN) to SVM for improving the performance of SVM near hyperplane because SVM concentrates on the macro-strategy of separating quasars from stars and kNN focuses on the special cases.

In this work, we focus on making a combined classifier of SVM and kNN to select the quasar candidates. The SDSS DR7 [14] contains more than 120000 identified quasars as our quasar sample to test the new method SVM-kNN. This paper is organized as follows. In sect. 2, we present the basic principles of SVM and kNN and describe the construction of SVM-kNN. Sect. 3 describes the characteristics of data used in this experiment in detail. Sect. 4 discusses the performance of SVM-kNN for separating quasars from stars. In sect. 5, we give our conclusion about SVM-kNN and what should be improved in the future work.

2 Methods

2.1 SVM

Support Vector Machines, proposed by Vapnik [15], is derived from the theory of structural risk minimization which belongs to statistical learning theory. The core idea of SVM is to map input vectors into a high-dimensional feature space and construct the optimal separating hyperplane in this space. SVM aims at minimizing an upper bound of the generalization error through maximizing the margin between the separating hyperplane and the data.

For a given training set belonging to two different classes,

$$T=(\boldsymbol{x}_1,\boldsymbol{y}_1),\cdots,(\boldsymbol{x}_n,\boldsymbol{y}_n),\ \boldsymbol{x}_s\in\mathbb{R}^{\mathrm{N}},y_s\in\{-1,+1\}. \tag{1}$$

SVM learns linear threshold functions of the type

$$h(\boldsymbol{x}_s)=\ \mathrm{sign}\{\boldsymbol{\omega}\cdot\boldsymbol{x}_s+b\}=\begin{cases}+1, & \text{if } \boldsymbol{\omega}\cdot\boldsymbol{x}_s+b>0,\\ -1, & \text{else.}\end{cases} \tag{2}$$

Each linear threshold function corresponds to a hyperplane in feature space and the side of the hyperplane on which an example $\boldsymbol{x}_s$ lies determines the calssified result by the function $h(\boldsymbol{x}_s)$. If the training data can be separated by at least one hyperplane h', the optimal hyperplane with maximum margin can be found by minimizing

$$F(\boldsymbol{\omega},\boldsymbol{\varepsilon})=\frac{1}{2}(\boldsymbol{\omega},\boldsymbol{\omega})+C\sum_{i=1}^{n}\varepsilon_s, \tag{3}$$

which subjects to

$$y_s[\boldsymbol{\omega}\cdot\boldsymbol{x}_s+b]\geqslant 1-\varepsilon_s,\ s=1,\cdots,n, \tag{4}$$

$$\varepsilon_s>0,\ s=1,\cdots,n. \tag{5}$$

The factor C is used to trade off training error and model complexity and ε are slack variables responding to the wrong prediction. The books [15,16] contain good description of SVM and Burges [17] provides good background material.

2.2 kNN

In pattern recognition, the kNN algorithm (kNN: Dudani [18], Beyer et al. [19]) is a simple method for classifying objects based on closest training examples in the feature space. The solution is defined as follows: Given a collection of data points and a query point in an *m*-dimensional feature space, find the data point closest to the query point. It is a type of instance-based learning, or lazy learning algorithm, since no explicit training step is required. In the classification phase of kNN, the value of *k* is a user-defined constant, and an unlabeled celestial object is classified by assigning the label which is most frequent among the *k* training samples nearest to that query celestial object. Euclidean distance is usually used as the distance metric. The kNN approach has many applications in astronomy, for example, Li et al. [20] had successfully applied kNN on classification of celestial objects.

2.3 The combined algorithm

We combine SVM and kNN together to generate a whole new classifier SVM-kNN. In order to keep the generalization capability of SVM, we firstly use SVM to preselect quasar candidates and then input its result into kNN to make a further selection. A source, for example, is firstly marked

as quasar by SVM. If over two thirds of its neighbors were stars in the feature space of kNN, this source will be modified as star by kNN. This strategy is not the same as the typical kNN method which gives an unknown source the category label most common amongst its kNN. In this study, we set k = 9 and only when the number of the nearest neighbors with oppositive category reaches 7, 8 or 9 can kNN change the category predicted by SVM. The basic idea is that we always believe the label of sample predicted by SVM, and only when the result of kNN is strongly opposed can we change the label. Therefore, the result of SVM-kNN is not the intersection of the two results of using SVM and kNN to select quasar candidates separately but the outcome of kNN based on the prediction of SVM. This new method has three new advantages:

(1) SVM-kNN can keep the generalization capability of SVM and prevent the over-fitting problem of kNN.

(2) SVM-kNN can correct the wrong predictions of some samples which are near the optimal separating hyperplane in SVM with the help of kNN.

(3) SVM-kNN can give an approximate probability for every quasar candidate using the number of nearest neighbors.

The generalization capability and the stability of SVM is a good candidate to make initial judgment for separating quasars from stars. It outperforms kNN when the number of positive training instances is small or there is a data skew problem. When facing the testing of out-of-sample, it can also maintain a relatively high performance. Nevertheless, kNN often obtains lower performance when the completeness of training samples is not sufficient. Actually, the basic "majority voting" classification causes that the more frequent examples tend to dominate the prediction of kNN, as they tend to come up in the kNN when the neighbors are computed due to their large number. Although SVM is robust, it is difficult to distinguish the type of samples which are located near the optimal separating hyperplane. Thus these samples are more likely to be mistaken only with SVM. For this reason, we add kNN to SVM for amending these mistakes made by SVM. In addition, the votes of a sample's neighbors can be thought of as the prediction probability of this sample and nevertheless this is a weakness of SVM which cannot give the degree of classification confidence.

2.4 Performance measurement

Besides the overall classification accuracy, we use metrics such as true negative rate, true positive rate, G-mean (GM), precision, recall, and F-measure (FM) to evaluate the performance of classification algorithms [21]. These metrics have been widely used for comparison of different classifiers. All these metrics are functions of the confusion matrix as shown in Table 1. TP is short for the true positive, FN for the false negative, FP for the false positive, TN for the true negative. In the process of classification, quasars are labeled as positive, stars as negative. The rows of the matrix are actual classes, and the columns are the predicted classes. Based on Table 1, the above-mentioned metrics are defined as follows:

$$\text{Accuracy}\left(\text{Acc}\right)=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}}, \tag{6}$$

$$\text{True Positive Rate}\left(\text{Acc}^{+}\right)=\frac{\text{TP}}{\text{TP}+\text{FN}}=\text{Recall}, \tag{7}$$

$$\text{True Negative Rate}\ \left(\text{Acc}^{-}\right)=\frac{\text{TN}}{\text{TN}+\text{FP}}, \tag{8}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}}, \tag{9}$$

$$\text{F-measure}\left(\text{FM}\right)=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}, \tag{10}$$

$$\text{G-mean}\left(\text{GM}\right)=(\text{Acc}^{-}\times\text{Acc}^{+})^{\frac{1}{2}}. \tag{11}$$

Recall is the fraction of actual positive cases that were correct, and precision is the fraction of the predicted positive cases that were correctly identified. For any classifier, there is always a tradeoff between recall and precision. The Geometric Mean (G-mean) is useful to determine "average factors". The F-measure can be interpreted as a weighted average of the precision and recall. These metrics are commonly used in the information retrieval area as performance measures. We will adopt all these measurements to compare our methods with different patterns. Train-test and ten-fold cross-validation were carried out to obtain all the performance metrics.

3 The data

Because most quasars fall into the star-like category from images, we extract point sources (type=6) with i-band magnitudes between 14.5 and (de-reddened) 21.3 [9] with PhotoPrimary view by specifically using the SQL interface to Catalog Archive Server (CAS). For the training sample in our study, we select the examples from the SpecPhotoAll table which combines spectral and photometric information of objects with specClass=1,3,4,6 (specClass=1 stars; specClass=3 quasars; specClass=4 high-redshift quasars (z > 2.3); specClass=6 late-type stars) and psf magnitude> 0

Table 1 Confusion matrix

	Predicted positive class	Predicted negative class
Actual positive class	TP (True Positive)	FN (False Negative)
Actual negative class	FP (False Positive)	TN (True Negative)

(rejecting object's magnitude=–9999). In addition, we give another constraint condition on samples with the psfMagErr< 0.1 for every bandpass since we want to understand how magnitude error can affect the performance of SVM-kNN (Large magnitude error will make the color error calculated by two magnitudes become larger). The sample is broken into two parts: one marked as '+1'(QSO) contains quasars and high-redshift quasars and the other labelled as '−1'(S) contains stars and late-type stars. In this work, the magnitudes all refer to psf magnitudes.

In Table 2, we list the total number of the four celestial object types under different selection conditions. In the third column, the sample of S is almost four times larger than QSO and this means that selecting quasar candidates is an imbalanced data problem. Although the SpecPhotoAll table is a precomputed join between the PhotoObjAll and SpecObjAll tables, a few objects in the PhotoObjAll table do not have the data of magnitude. Therefore the celestial object counts with psfMag>0 are slight decrease in the second column. In the first column, it shows that most high-redshift quasars and late-type stars are removed out of samples under the limitation of psfMagErr< 0.1 and the ratio of S to QSO turns 3.

In order to understand the influence of the limitation of psfMagErr< 0.1 on the four spectral types, the magnitude as a function of magnitude error and the magnitude distribution are plotted. Figure 1 indicates that magnitude errors grow as magnitudes increase and the accuracy of u-band magnitude is usually lower than those in the other four bands for these four spectral types. It also indicates that the magnitude of many high-redshift quasars exceeds the limiting magnitude in u band but late-type stars do in all bands. Figure 2 further describes the magnitude distribution of the four spectral types. In the bottom two panels, it is shown that the gray areas with psfMagErr> 0.1 are larger for high-redshift quasars and late-type stars. Thus many high-redshift quasars and late-type stars are removed through magnitude error limitation.

4 Quasar selection

SVM-kNN is based on SVM and kNN so we have to preset model parameters for them. Peng et al. [22] describe that how to tune the model parameters of SVM in detail. According to that experiment, we give one optimal combination ($C_{-+} = 1$, $C_{+-} = 10$, kernel = RBF, $\gamma = 1$) for using SVM to select quasar candidates in SDSS DR7. The experimental results of that work demonstrated that the precision and the recall of SVM for separating quasars from stars both can be high. For kNN classification, there are two important model parameters (distance and the number of nearest neighbors) to be set. We use empirical distance= 'Euclidean' and nearest neighbors $k = 9$ as the model parameters. Actually, the influence of distance type and the k value is rather small and there is a detailed discussion in ref. [23]. For the two selection steps in SVM-kNN, we input two different input patterns. The best input pattern of SVM is (u-g, g-r, r-i, i-z, z) and the best input pattern of kNN is (u-g, g-r, r-i, i-z) with the optical photometry in SDSS data. Here u, g, r, i, z all point to dereddened SDSS psfMag u, g, r, i, z according to the dust reddening map of Schlegel et al. [24], respectively. Therefore, there are two kinds of feature space for SVM-kNN.

SVM can use a relatively small sample for training because of its good generalizations so we randomly divide the sample sets of confirmed celestial objects into two parts, the training set and the testing set. The ratio between them is 1–9. Firstly, we evaluate the performance of SVM-kNN for separating quasars from stars without restriction on magnitude error. Table 3 shows that accuracy, precision, recall and G-mean have been improved by adding kNN into SVM. Improving magnitude quality brings much higher performance because the magnitudes with smaller errors get more reliable colors and SVM-kNN does not offer a solution for including magnitude error at present. After restricting magnitude error less than 0.1, the recall (quasars) rises because the confusing objects which usually contain high errors has been filtered out. In order to carefully study the performance of SVM-kNN, Tables 4 and 5 provide classification results. They clearly show that there is a challenge for separating high redshift quasars from stars and late type stars, and the recall of high redshift quasars reaches only about 80%. Even though SVM-kNN selects quasars from stars using high-quality data in Table 5, the recall of high redshift quasars is still inadequate. The main reason is that the number of stars and quasars is far more than that of high redshift quasars and most support vectors on the hyperplane of SVM come from the sample set of stars and quasars.

5 Conclusion

In this paper, we slightly improve the performances of SVM

Table 2 Sample from the SDSS DR7 SpecPhotoAll table

SpecClass	psfMag > 0 & psfMagErr < 0.1 (No.)	psfMag > 0 (No.)	Original data (No.)
Stars(1)	289468	366567	367806
Quasars(3)	94893	110221	110453
High-redshift quasars(4)	4507	9491	9507
Late-type stars(6)	8883	79934	80147
Total	397751	566213	567913

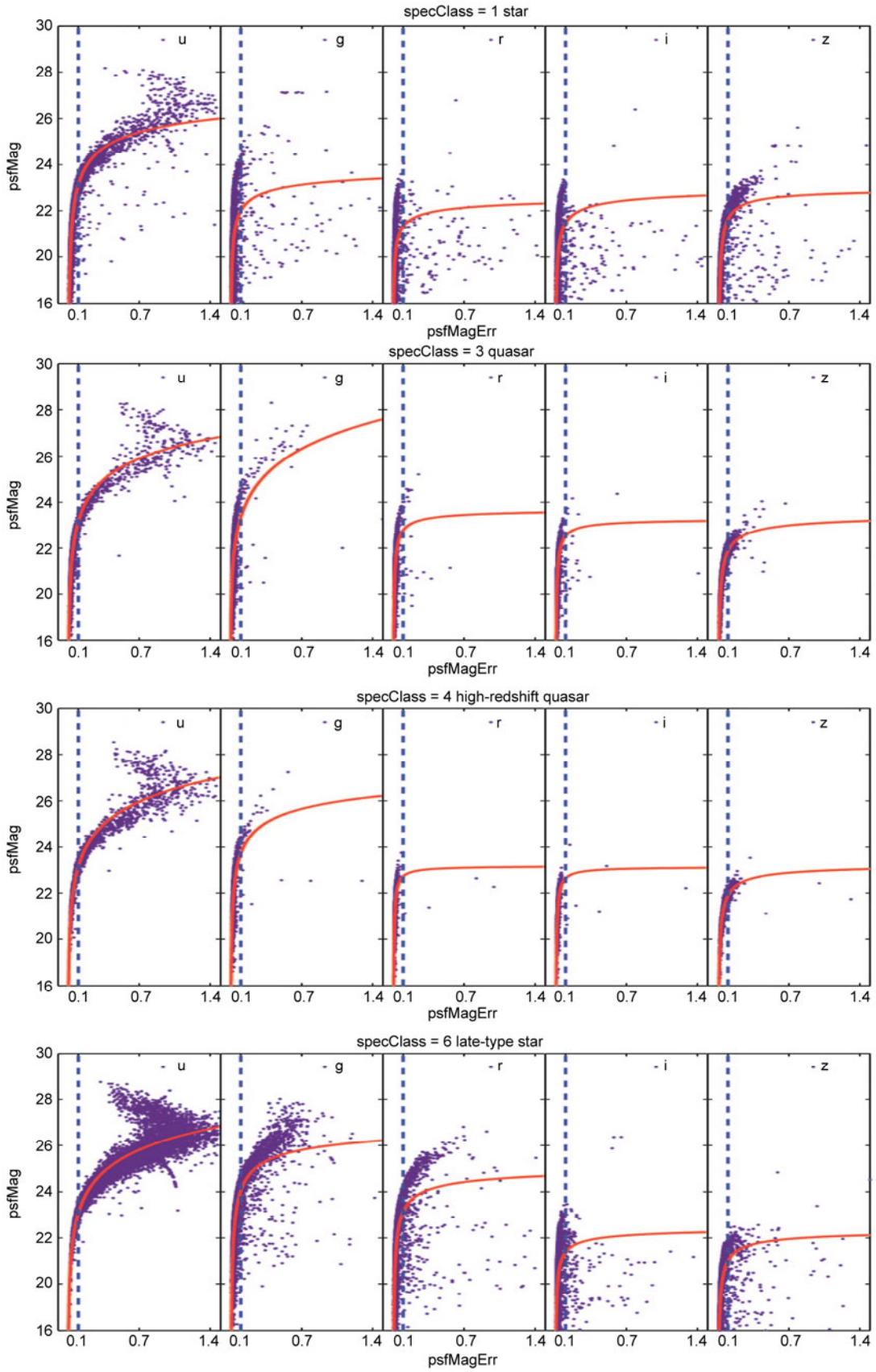

Figure 1 (Color online) Scatter of psfMag error of the four spectral types of objects. The dashed lines indicate the limitation of magnitude error. The solid line is the curve fitting of the points of magnitude error.

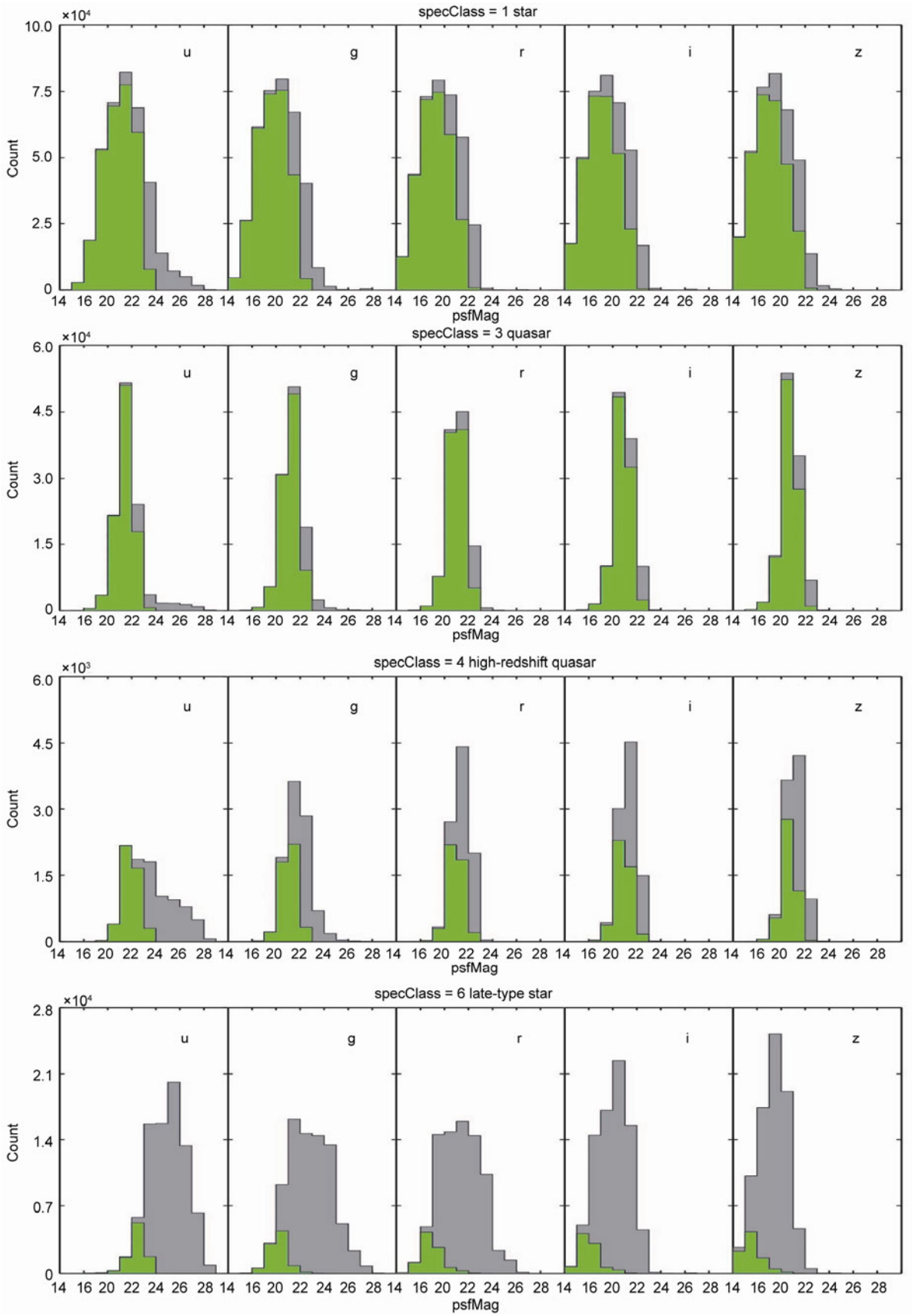

Figure 2 Magnitude distribution of the four spectral types of objects. The green area represents the celestial objects satisfied with psfMagErr< 0.1 and psfMag > 0 and the gray area indicates the celestial objects satisfied with psfMag > 0.

Table 3 Performance comparison of different methods

Method	Accuracy (%)	Precision (%)	Recall (%)	G-mean (%)
SVM	97.89	95.93	93.94	94.97
SVM-kNN	97.99	96.48	94.01	95.23
SVM(MagErr<0.1)	98.68	97.56	97.13	97.34
SVM-kNN(MagErr<0.1)	98.85	98.08	97.30	97.69

Table 4 Classification result of SVM-kNN without the restriction of magnitude error

SpecClass	Stars	Quasars	High-*Z* quasars	Late-type stars
No. total	329903	99201	8540	71948
No. rightly classified	326576	94314	6900	71578
No. misclassified	3327	4887	1640	370
Accuracy	98.99%	95.07%	80.80%	99.49%

Table 5 Classification result of SVM-kNN with magnitude error<0.1

SpecClass	Stars	Quasars	High-*Z* quasars	Late-type stars
No. total	260540	85414	4046	7976
No. rightly classified	259011	83865	3179	7870
No. misclassified	1529	1549	867	169
Accuracy	99.41%	98.19%	78.57%	97.88%

to select quasar candidates through combining SVM and kNN and create a new method named SVM-kNN. The generalization of SVM (the ratio of training set to testing set is 1:9) is retained in SVM-kNN and the performances of it is raised by using kNN as a useful complement. In addition, the result of SVM-kNN shows that if we consider the magnitude error limitation of objects, this approach can obtain good performances, for example, the precision and the recall of quasars are both above 97.0%. Nevertheless, the low performance of high-redshift quasars should be raised in the future work with the number of this sample increasing. There are three potential solutions: a) using multi-band photometry data; b) improving the percentage of the population of high- redshift quasars; c) introducing a new method into SVM-kNN to classify objects with large magnitude errors. In brief, this SVM-kNN approach is promising to be applied to preselect quasar candidates from various photometric databases for large spectroscopic sky survey projects.

We are very grateful to the constructive comments and suggestions made by the referees. This work was supported by the National Natural Science Foundation of China (Grant Nos. 10778724, 11178021 and 11033001), the Natural Science Foundation of Education Department of Hebei Province (Grant No. ZD2010127) and the Young Researcher Grant of National Astronomical Observatories, Chinese Academy of Sciences. We acknowledgment SDSS database. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

1 Tyson J A. Large synoptic survey telescope: Overview. Soc Photo-Opt Instrum Eng Conf Ser, 2002, 4836: 10–20
2 McPherson A M, Born A, Sutherland W, et al. VISTA: Project status. Soc Photo-Opt Instrum Eng Conf Ser, 2006, 6267: 7–20
3 Kaiser N, Aussel H. Pan-STARRS: A large synoptic survey telescope array. Soc Photo-Opt Instrum Eng Conf Ser, 2002, 4836: 154–164
4 Abraham S, Philip N S, Kembhavi A, et al. Photometric catalogue of quasars and other point sources in the sloan digital sky survey. Mon Not R Astron Soc, 2012, 419: 80–94
5 Carballo R, Gonźalez-Serrano J I, Benn C R, et al. Use of neural networks for the identification of new z >= 3.6 QSOs from FIRST-SDSS DR5. Mon Not R Astron Soc, 2008, 391: 369–382
6 Zhang Y X, Zhao Y H. Automated clustering algorithms for classification of astronomical objects. Astron Astophys, 2004, 422: 1113–1121
7 Richards G T, Fan X. Spectroscopic target selection in the sloan digital sky survey: The quasar sample. Astron J, 2002, 123: 2945–2975
8 Richards G T, Nichol R C, Gray A G. Efficient photometric selection of quasars from the sloan digital sky survey: 100,000 $z < 3$ quarsars from data release one. Astrophys J Suppl Ser, 2004, 155: 257–269
9 Richards G T, Myers A D, Gray A G. Efficient photometric selection of quasars from the sloan digital sky survey. II. 1, 000, 000 quasars from data release 6. Astrophys J Suppl Ser, 2009, 180(1): 67–83
10 Richards G T, Deo R P, Lacy M, et al. Eight-dimensional mid-infrared/optical bayesian quasar selection. Astron J, 2009, 137: 3884–3899
11 Gao D, Zhang Y X, Zhao Y H. Support vector machines and KD-tree for separating quasars from large lurvey data bases. Mon Not R As-

tron Soc, 2008, 386, 1417–1425
12 Bailer-Jones C A L, Smith K W, Tiede C, et al. Finding rare objects and building pure samples: Probabilistic quasar classification from low-resolution Gaia spectra. Mon Not R Astron Soc, 2008, 391: 1838–1853
13 Kim D W, Protopapas P, Byun Y I, et al. QSO selection algorithm using time variability and machine learning: Selection of 1,620 QSO candidates from MACHO LMC database. Astrophys J, 2011, 735: 68–84
14 Abazajian K N, Adelman-McCarthy J K, Agüeros M A, et al. The seventh data release of the sloan digital sky survey. Astrophys J Suppl Ser, 2009, 182(2): 543–558
15 Vapnik V. The Nature of Statistical Learning Theory. New York: Springer, 1995
16 Vapnik V. Statistical Learning Theory. New York: John Wiley and Sons, Inc., 1998
17 Burges C. A tutorial on support vector machines for pattern recognition. Pattern Recogn, 1998, 167: 121–167
18 Dudani S. The distance-weighted *k*-nearest-neighbor rule, systems, man and cybernetics. IEEE, 1976, 4: 325–327
19 Beyer K, Goldstein J, Ramakrishnan R, et al. When is nearest neighbor meaningful? Database Theory ICDT99, 1999, 217–235
20 Li L, Zhang Y, Zhao Y. *k*-Nearest Neighbors for automated classification of celestial objects. Sci China Ser G-Phys Mech Astron, 2008, 51: 916–922
21 Chen C, Liaw A, Breiman L. Using Random Forest to Learn Imbalanced Data. Technical Report 666, Department of Stastics, UC Berkeley. 2004, 1–12
22 Peng N, Zhang Y, Zhao Y. Support vector machines for quasar selection. Soc Photo-Opt Instrum Eng Conf Ser, 2010, 7740: 77402T
23 Peng N, Zhang Y, Zhao Y. Comparison of several algorithms for celestial object classification. Soc Photo-Opt Instrum Eng Conf Ser, 2010, 7740: 77402M
24 Schlegel D J, Finkbeiner D P, Davis M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. Astrophys J, 1998, 500: 525–553