Chin. J. Astron. Astrophys. Vol. 8 (2008), No. 1, 119–126
(*http://www.chjaa.org*)

Chinese Journal of
Astronomy and
Astrophysics

Two Novel Approaches for Photometric Redshift Estimation based on SDSS and 2MASS *

Dan Wang[1,2], Yan-Xia Zhang[1], Chao Liu[1,2] and Yong-Heng Zhao[1]

[1] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012;
dwang@lamost.org
[2] Graduate School of Chinese Academy of Sciences, Beijing 100049

Received 2007 March 28; accepted 2007 May 12

Abstract We investigate two training-set methods: support vector machines (SVMs) and Kernel Regression (KR) for photometric redshift estimation with the data from the databases of Sloan Digital Sky Survey Data Release 5 and Two Micron All Sky Survey. We probe the performances of SVMs and KR for different input patterns. Our experiments show that with more parameters considered, the accuracy does not always increase, and only when appropriate parameters are chosen, the accuracy can improve. For different approaches, the best input pattern is different. With different parameters as input, the optimal bandwidth is dissimilar for KR. The rms errors of photometric redshifts based on SVM and KR methods are less than 0.03 and 0.02, respectively. Strengths and weaknesses of the two approaches are summarized. Compared to other methods of estimating photometric redshifts, they show their superiorities, especially KR, in terms of accuracy.

Key words: galaxies: distances and redshifts — galaxies: general — methods: data analysis — techniques: photometric

1 INTRODUCTION

Photometric redshifts have been regarded as the most promising tool in the study of the formation and evolution of galaxies and the large scale structure of the universe, considering that the spectra of faint objects are difficult to obtain. The photometric redshift technique translates observable signals such as flux and apparent color to the corresponding intrinsic properties of absolute luminosity and rest-frame color. One purpose behind the photometric redshift technique is to measure the redshifts of galaxies and AGN based on multi-wavelength photometry. The photometric redshift technique can be traced back to Baum (1962) who used nine medium-wide filters to detect the 4000 Å in galaxies. For example, the predicted redshift of the C10925 galaxies by this technique is $z = 0.19$, which agrees closely with the known spectroscopic value of $z = 0.192$. Subsequent implementation has been made by Koo (1985) using four broad-band photographic filters, by Loh & Spillar (1986) using CCDs along with six medium-band filters, and by Xia et al. (2002) using CCD photometry of BATC 15 medium-band filters. In the last two decades, some well-defined statistical techniques have become increasingly popular in predicting photometric redshifts.

There are two approaches to estimate photometric redshift in the literature: template fitting, with templates derived from synthetic (e.g. Bruzual & Charlot 1993) or empirical spectra (e.g. Coleman, Wu & Weedman 1980), and empirical training set, which constructs a direct empirical correlation between color and redshifts. For the template fitting, some templates are constructed in advance according to the known redshifts and galaxy types. By minimizing the standard χ^2 to fit the observed photometric data with a set

* Supported by the National Natural Science Foundation of China.

of spectral templates, this method can be applied beyond the redshift limit. Although it is easy to carry out, the accuracy of this approach strongly depends on the templates. The training set approach, on the other hand, derives a functional relation between redshift and photometric data using a large and representative training set of galaxies with known photometry and redshifts. Then the functional relation is applied to estimate the redshifts of objects with unknown redshifts. In the last few years, a large number of training set methods has been developed and implemented (Way & Srivastava 2006). For example, linear or non-linear fitting (Brunner et al. 1997; Wang, Bahcall & Turner 1998; Budavari et al. 2005); support vector machines (SVMs, Wadadekar 2005); artificial neural network (ANNs, Firth, Lahav & Somerville 2003; Ball et al. 2004; Collister & Lahav 2004; Vanzella et al. 2004; Li et al. 2006).

Another training-set approach is the instance-based learning technique for predicting the photometric redshifts (e.g. Csabai et al. 2003; Ball et al. 2007), which needs no training, but implements the predictions directly on the data that have been stored in the memory. In general, they store all the training data in the memory during the learning phase, and defer all the essential computation until the prediction phase. Such techniques include the k-nearest neighbor, kernel regression and locally weighted regression.

In this paper we further explore two approaches: support vector machines (SVMs) and kernel regression (KR), with a view of estimating the redshifts of galaxies with photometric data from the SDSS and 2MASS databases. The structure of this paper is as follows: Section 2 describes the data used. Section 3 describes the principles of SVMs and KR. Section 4 gives the results and a discussion. Finally conclusions are summarized in Section 5.

2 DATA

The data used in this paper are from the Sloan Digital Sky Survey (SDSS) and the Two Micron All Sky Survey (2MASS). Some general information on these are as follows.

The Sloan Digital Sky Survey (SDSS, York et al. 2000) is an astronomical survey project, which covers more than a quarter of the sky, to construct the first comprehensive digital map of the universe in 3D using a dedicated 2.5-meter telescope located at Apache Point, New Mexico. In its first phase of operations, it has imaged 8,000 square degrees in five bandpasses (u, g, r, i, z) and measured more than 675,000 galaxies, 90,000 quasars and 185,000 stars. In its second stage, SDSS will carry out three new surveys in different research areas, such as the nature of the universe, the origin of galaxies and quasars, and the formation and evolution of the Milky Way.

The Two Micron All Sky Survey (2MASS, Cutri et al. 2003) uses two highly-automated 1.3-m telescopes, one at Mt. Hopkins, Arizona, the other at CTIO, Chile. Each telescope is equipped with a three-channel camera, and each channel consists of a 256×256 array of HgCdTe detectors, capable of observing the sky simultaneously at J (1.25 μm), H (1.65 μm), and Ks (2.17 μm), to a 3σ limiting sensitivity of 17.1, 16.4 and 15.3 mag in the three bands, respectively. Jarrett et al. (2000) had given more detailed information in the extended source catalog.

We select all galaxies with known spectral redshifts from SDSS Data Release 5, and then cross-match the data with 2MASS extended source catalog within a search radius of 3 times the SDSS positional errors. The cross-matching generated about 150,000 galaxies. From these we select objects satisfying the following criteria: 1) the spectroscopic redshift confidence must be equal to or greater than 0.95; 2) redshift warning flag is 0; 3) $r < 17.5$. This resulted in a sample of 62,083 galaxies. Table 1 describes the broadband filters and their wavelength range from SDSS and 2MASS catalogs.

3 MODEL SELECTION

3.1 Support Vector Machines

The primary conception of SVMs was developed by Vapnik (1995). SVMs were developed to solve the classification problem, but recently they have been extended to the domain of regression. Regression of SVMs is achieved by using an alternative loss function, which must be modified to include a distance measure. The SVM task usually involves training and testing data which consist of some data instances. Each instance in the training set contains one "target value" and several "attributes". The goal of SVMs is to produce a model which predicts the target value of data instances in the testing set when only the attributes are given.

Table 1 Survey Filters and Characteristics

Bandpass	Survey	λ_{eff} (Å)	$\Delta\lambda$ (Å)
u	SDSS	3551	600
g	SDSS	4686	1400
r	SDSS	6165	1400
i	SDSS	7481	1500
z	SDSS	8931	1200
J	2MASS	12500	1620
H	2MASS	16500	2510
Ks	2MASS	21700	2620

Given a training set of training pairs (x_1, y_1),..., (x_l, y_l), $x_i \in R^n$, $y \in R$, with a linear function,

$$f(x) = \langle \omega, x \rangle + b. \tag{1}$$

The optimal regression function is given by the minimum of the functional,

$$\phi(\omega, \zeta) = \frac{1}{2}\omega.\omega + C\sum_i (\zeta_i^- + \zeta_i^+). \tag{2}$$

Using a quadratic loss function,

$$L_{\text{quad}}(f(x) - y) = (f(x) - y)^2, \tag{3}$$

the solution is given by,

$$\max_{\alpha,\alpha^*} W(\alpha, \alpha^*) = \max_{\alpha,\alpha^*} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j \rangle$$

$$+\sum_{i=1}^{l}(\alpha_i - \alpha_i^*)y_i - \frac{1}{2C}\sum_{i=1}^{l}(\alpha_i^2 + (\alpha_i^2)^2), \tag{4}$$

and the resultant optimization is

$$\min_{\beta} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\beta_i\beta_j\langle x_i, x_i \rangle - \sum_{j=1}^{l}\beta_i y_i + \frac{1}{2C}\sum_{i=1}^{l}\beta_i^2, \tag{5}$$

with constraint

$$\sum_{i=1}^{l}\beta_i = 0. \tag{6}$$

To generalize to non-linear regression, we replace the dot product with a kernel function. More information can be found in Steve's tutorial (1998). In this work we adopt the Gaussian kernel function.

SVMs have been widely used in the area of machine learning because of its excellent generalization performance, such as handwritten digit recognition and face detection. In astronomy, SVMs have been applied to identifying red variables (Williams et al. 2004), clustering of astronomical objects (Zhang & Zhao 2004), and classifying AGN from stars and normal galaxies (Zhang, Cui & Zhao 2002).

Several software implementations of the SVM algorithm are accessible on the web. Considering the robustness, the ability of handling large amounts of data, and the regression time, we selected for our study SVM_Light, a fast optimized SVM algorithm, implemented in the C language. It can deal with many thousands of support vectors, handle hundreds/thousands of training examples, and can provide several standard kernel functions. The details about SVM_Light can be found at *http://www.cs.cornell.edu/People/tj/svm_light/*.

3.2 Kernel Regression

KR belongs to the family of instance-based learning algorithms (Watson 1964; Nadaraya 1964), which simply stores some or all of the training examples and does not perform any generalization of the given samples, and "delays the learning" till the prediction time. Given a query point $x_{\rm q}$, a prediction is obtained using the training samples that is "most similar" to $x_{\rm q}$. Subsequently, similarity is measured by means of a distance metric defined in the hyper-space of V predictor variables. KR gives the prediction for a query point $x_{\rm q}$, by a weighted average of the y values of its neighbors. The weight of each neighbor is calculated according to a function of its distance to $x_{\rm q}$ (the kernel function). These kernel functions give more weight to neighbors that are nearer to $x_{\rm q}$. The notion of neighborhood (or bandwidth) is defined in terms of the distance from $x_{\rm q}$. The prediction for query point $x_{\rm q}$ is obtained by

$$y_{\rm q} = \frac{\sum\limits_{i=1}^{N} K\left(\frac{D(x_i, x_{\rm q})}{h}\right) \times y_i}{\sum\limits_{i=1}^{N} K\left(\frac{D(x_i, x_{\rm q})}{h}\right)}, \tag{7}$$

where $D(.)$ is the distance function between two instances, $K(.)$ is a kernel function, h the bandwidth value, and (x_i, y_i) the training samples. In this paper, we use Euclidian distance and Gaussian kernel function. Here x_i is the feature for each training sample, y_i the spectroscopic redshift for each training set sample, and $y_{\rm q}$ the redshift of each query sample.

When using KR an important design decision is to select the bandwidth h. A larger h would result in a flatter weight function curve, indicating that many points of the training set contribute quite evenly to the regression. As the h tends to infinity, the predictions would approach the global average of all the points in the database. If h is very small, only the closely neighboring data points make a significant contribution. If the data are relatively noisy, we expect to obtain smaller prediction errors with a relatively larger h. If the data are noise-free, then a small h will avoid smoothing out the finer details. There exist well-tested algorithms for choosing the bandwidth for the KR which minimizes the differences between the true underlying distribution and the estimated distribution. Usually the selection of the bandwidth is done by cross-validation.

In cross-validation, we first divide the given sample into subsets, then perform a preliminary analysis on one such subset, and use the other subsets for confirming and validating the initial analysis. In an M-fold cross-validation, the data are first divided into m subsets of approximately equal size. Then each of the m subsets is used in turn as the test set and the other $m-1$ subsets are put together to form a training set for a given bandwidth. Then, the average error across all the m trials is computed (Zhang & Zhao 2007). Here we adopt 10-fold cross-validation for the bandwidth choice, dividing the samples into ten subsets. The optical bandwidth is indicated by the bandwidth with the minimum average errors. In Table 2 we apply KR with seven-color ($u-g, g-r, r-i, i-z, z-J$, $J-H$ and $H-Ks$) and spectra redshifts as an input pattern, taking it as an example to illustrate the relationship between bandwidth (h) and cross-validated value (CV). It is found that the optimal bandwidth h is 0.045 when cross-validated value arrives at the minimum value 4.33.

Table 2 The Relationship between Bandwidth (h) and Cross-Validated Value (CV)

h	0.015	0.02	0.025	0.03	0.035	0.04	**0.045**	0.05	0.055	0.06
CV($\times 10^{-5}$)	4.77	5.03	4.86	4.64	4.45	4.35	**4.33**	4.35	4.41	4.77

4 RESULT AND DISCUSSION

One advantage of the empirical training set approach to photometric redshift estimation is that additional parameters can be easily incorporated. Additional parameters (e.g. $petro50\_r$, $petro90\_r$, and $fracDeV\_r$, etc.) may be added to the input. We exmained different input patterns in order to study which parameter influences the accuracy of the predicted photometric redshifts. We randomly divide the sample into two parts: 41,388 for the training and 20,695 for the testing, and applied KR and SVM for various sets of input parameters. The resulting rms deviations are listed in Table 3.

Using SVMs to estimate the photometric redshifts, best performance is achieved with input sets of colors. The 4-color input ($u-g, g-r, r-i$ and $i-z$) gives the same best accuracy ($\sigma_{\rm rms}$=0.0273) as that based on the 7-color input ($u-g, g-r, r-i, i-z, z-J, J-H$ and $H-Ks$). The set of seven colors plus the r magnitude is better than the set of four colors plus r magnitude, which is better than the set of seven magnitudes. The performance taking five magnitudes as input is not as good as taking seven magnitudes. Since the accuracy with four colors is best, we considered adding more parameters to see whether the performance improves even more. As shown in Table 3, the accuracy decreases when adding $fracDev\_r$ or $petro50\_r$ and $petro90\_r$. Thus, adding more parameters does not always improve the performance, sometimes even makes it worse. The results have shown that there is no significant improvement when we take more parametric data from the SDSS and 2MASS catalogs, i.e., it increases the number of attributes markedly but does not decrease the rms error. So this course is not recommended.

Table 3 Dispersions of Photometric Redshift Prediction Using KR and SVMs

Method	KR	SVMs
Input Parameters	$\sigma_{\rm rms}$(optimal bandwidth)	$\sigma_{\rm rms}$
u, g, r, i, z	0.0208 ($h = 0.025$)	0.0291
u, g, r, i, z, J, H, Ks	0.0254 ($h = 0.015$)	0.0278
$u-g, g-r, r-i, i-z$	0.0193 ($h = 0.020$)	0.0273
$u-g, g-r, r-i, i-z, r$	0.0196 ($h = 0.025$)	0.0284
$u-g, g-r, r-i, i-z, z-J, J-H, H-Ks$	0.0210 ($h = 0.045$)	0.0273
$u-g, g-r, r-i, i-z, z-J, J-H, H-Ks, r$	0.0235 ($h = 0.055$)	0.0275
$u-g, g-r, r-i, i-z, fracDev\_r$	0.0192 ($h = 0.020$)	0.0306
$u-g, g-r, r-i, i-z, petro50\_r, petro90\_r$	0.0218 ($h = 0.040$)	0.0330

Note — $petro50\_r$ is the Petrosian 50% radius in r band, $petro90\_r$ the Petrosian 90% radius in r band, and $fracDeV\_r$ is $fracDeV$ in r band.

For KR, the best input pattern is the four colors ($u-g$, $g-r$, $r-i$ and $i-z$) plus $fracDev\_r$ when the rms error amounts to 0.0192. The next best patterns are the four colors by themselves or four colors plus the r magnitude, when rms error is 0.0193 or 0.0196, respectively. Then comes the input set of the five magnitudes when the rms scatter is 0.0208. The result with just seven colors is better than that with seven colors plus with the r magnitude, but worse than that of five magnitudes. For four colors as inputs, the performance of KR is made worse when adding $petro50\_r$ and $petro90\_r$, except $fracDev\_r$. Thus, when using KR to predict photometric redshifts, we find the parameters other than the magnitudes and color indices, such as $petro50\_r$ and $petro90\_r$, are ineffective; however, $fracDev\_r$ is important and effective, possibly because $fracDev\_r$ is closely related to galaxy type. When implementing KR, an increasing bandwidth may cause a small loss of estimation. In our experiments, the fraction of loss is less than 1%. Table 3 also indicates that the optimal bandwidth is different for different input patterns.

Figure 1 (2) plots the photometric redshifts calculated with KR (SVMs) against the known spectroscopic redshifts. The left (right) panels correspond to the best (worst) input set. It is clear that the SVM-estimated photometric redshifts are too high for low redshifts, but the KR-estimated photometric redshifts are rather satisfactory.

To compare the performance of different methods for photometric redshift estimation, we list the rms scatters of photometric redshifts from the different studies in Table 4. Because the accuracy strongly depends on the data, the comparison is approximate. As shown in Table 4, the kernel regression (KR) is comparable to the artificial neural networks (ANNs), which is better than the SVMs (Wadadekar 2005), the Kd-tree (Csabai et al. 2003), the polynomial (Connolly et al. 1995), and superior to CWW and Bruzual-Charlot (Csabai et al. 2003). Nevertheless, each method has its strong and weak points. KR belongs to instance-based learning family. It is a form of memory-based method, only learning at the prediction phase, therefore, it takes a large memory of the computer, even though it is of high accuracy. With the ANNs, one should be familiar with the network architecture and require judicious decision as to the number of input

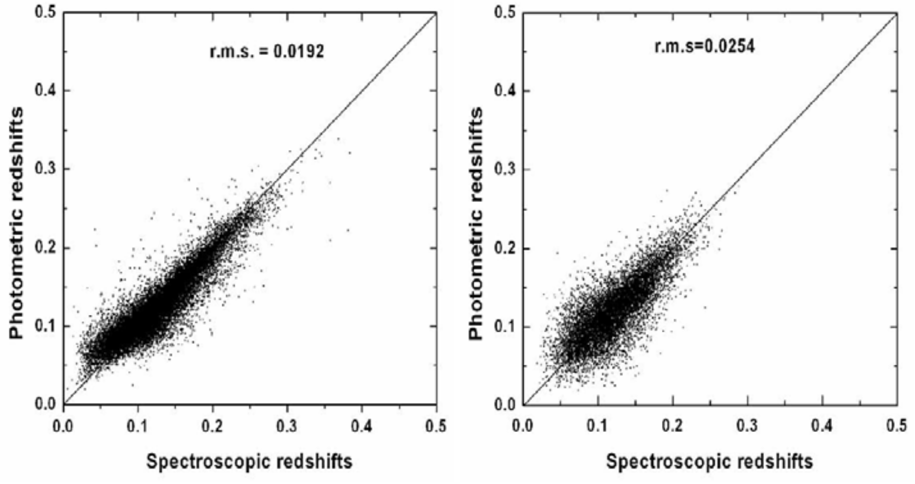

Fig. 1 KR-calculated photometric redshifts for 62,083 galaxies from the SDSS DR5 and 2MASS plotted against the known spectroscopic redshifts. Left: with the best input set ($u - g, g - r, r - i, i - z$ and $fracDev\_r$). Right: with the worst input set (u, g, r, i, z, J, H and Ks).

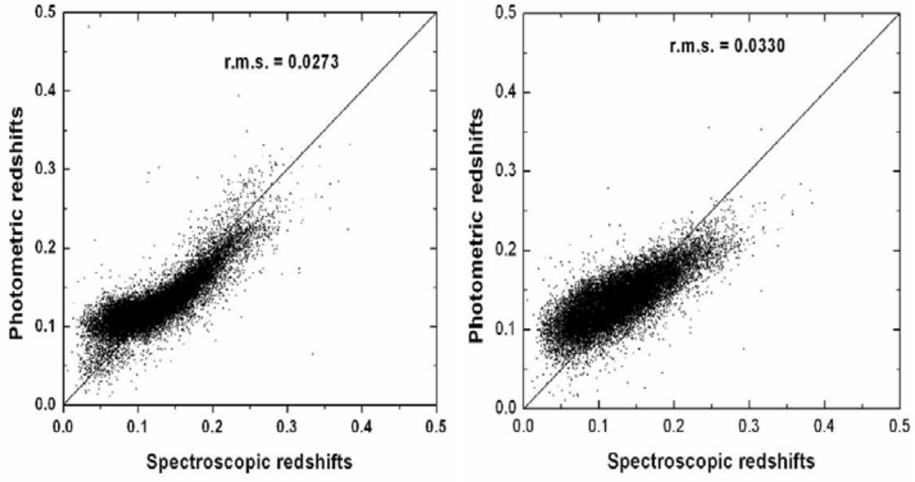

Fig. 2 SVM-calculated photometric redshifts for 62,083 galaxies from the SDSS DR5 and 2MASS plotted against the known spectroscopic redshifts. Left: with the best input set ($u - g, g - r, r - i$ and $i - z$). Right: with the worst input set ($u - g, g - r, r - i, i - z, petro50\_r$ and $petro90\_r$).

nodes and hidden layers: the more complex the networks, the more accurate the result is. However, SVMs may use different kernel functions instead of different architectures like ANNs. As soon as one chooses an appropriate kernel function and parameters, the rms scatter will decrease significantly. Moreover, such classical problems as multi-local minima, curse of dimensionality and overfitting in ANNs, seldom occur in SVMs. Nevertheless, SVMs need prior knowledge to adjust the parameters. Degeneration among the parameters complicates the process of regulation. Even though linear or non-linear polynomial regression is easy to implement and is well-known to astronomers, it involves large systematic deviations (Brunner et al. 1997; Wang et al. 1998; Budavári et al. 2005; Hsieh et al. 2005; Connolly et al. 1995). Csabai et al. (2000) have presented a hybrid model that is a combination of a template-based and an empirical training set. The hybrid model can reconstruct the continuum spectra of galaxies directly from a set of multicolor photometric observations and spectroscopic redshifts. Although using this hybrid model the dispersion of the photometric redshifts is significantly improved, it is still not as good as the empirical dispersion.

Table 4 Accuracy of Photometric Redshifts Derived from Various Approaches

Method Name	$\sigma_{\rm rms}$	Data set	Input parameters
CWW[1]	0.0666	SDSS-EDR	$ugriz$
Bruzual-Charlot[1]	0.0552	SDSS-EDR	$ugriz$
Interpolated[1]	0.0451	SDSS-EDR	$ugriz$
Polynomial[1]	0.0318	SDSS-EDR	$ugriz$
Kd-tree[1]	0.0254	SDSS-EDR	$ugriz$
ClassX[2]	0.0340	SDSS-DR2	$ugriz$
SVMs[3]	0.0270	SDSS-DR2	$ugriz$
ANNs[4]	0.0229	SDSS-DR1	$ugriz$
Polynomial[5]	0.0250	SDSS-DR1,GALEX	$ugriz + nuv$
KR	0.0208	SDSS-DR5,2MASS	$ugriz$
	0.0193	SDSS-DR5,2MASS	$color^*$
SVMs	0.0273	SDSS-DR5,2MASS	$color^*$

Note — SDSS-EDR = Early Data Release (Stoughton et al. 2002),
SDSS-DR1 = Data Release 1 (Abazajian et al. 2003),
SDSS-DR2 = Data Release 2 (Abazajian et al. 2004),
SDSS-DR5 = Data Release 5 (Adelman-McCarthy et al. 2007).
$color^*$ is the color indexes, i.e. $u - g$, $g - r$, $r - i$ and $i - z$.
(1) Csabai et al. (2003); (2) Suchkov, Hanisch & Margonet (2005);
(3) Wadadekar (2005); (4) Collister & Lahav (2004); (5) Budavári et al. (2005).

5 CONCLUSIONS

We use two novel methods, SVMs and KR, to estimate the photometric redshifts of objects common to SDSS DR5 and 2MASS. We compare their performance for various input sets. Our results show that only by choosing an appropriate input set of parameters can the accuracy be improved. Additional bandpasses from the infrared (2MASS) contributes little information due to the small size of data set, and there is no significant improvement by adding the morphological parameters ($petro50\_r, petro90\_r$ and $fracDev\_r$).

The accuracy of SVM-derived photometric redshifts is slightly less than that using ANNs, as good as using linear or quadratic regression, and clearly much better than using template fitting. In certain situations, SVMs will be a highly competitive tool for determining photometric redshifts with regard of speed and application. However, it does require a large and representative training sample. As an empirical estimator, it is impossible to extrapolate SVMs to regions not well sampled by the training set. Moreover, a potential way to increase the accuracy of the photometric redshifts is to choose a more appropriate kernel function, and to consider the methods of feature selection/extraction in the process of the input parameter selection.

The dispersion of photometric redshift estimation by KR is fairly satisfactory. Compared to other training-set methods, KR does not need any extra effort on the training. In addition, KR improves the empirical training-set methods. Even when the sample contains a few high-redshift galaxies, KR can appropriately adjust the bandwidth to obtain much more accurate redshifts. Thus, KR can be extrapolated to regions where the input parameters are not well represented by the training set. As larger and deeper photometric surveys are carried out, it seems that the KR will show its superiority. We plan to explore adaptive bandwidth or other kinds of distance metric for KR in a future study.

Acknowledgements The authors acknowledge an anonymous referee whose insightful and detailed comments has improved the presentation of this paper. This paper is funded by the National Natural Science Foundation of China under grants Nos.10473013, 10778724 and 90412016.

This research has made use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan

Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. 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, Cambridge University, 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.

References

Abazajian K. et al., 2003, AJ, 126, 2081
Abazajian K. et al., 2004, AJ, 128, 502
Adelman-McCarthy J. et al., 2007, ApJS, 172, 634
Baum W. A., 1962, IAU Symp. 15, 390
Ball N. M., Brunner R. J., Myers A. D. et al., 2007, ApJ, 663, 774
Ball N. M., Loveday J., Fukugita M. et al., 2004, MNRAS, 348, 1038
Brunner R. J., Connolly A. J., Szalay A. S. et al., 1997, ApJ, 482, 21
Bruzual A. G., Charlot S., 1993, ApJ, 405, 538
Budavari T. et al., 2005, ApJ, 619, 31
Coleman G. D., Wu C. C., Weedman D. W., 1980, ApJS, 43, 393
Collister A. A., Lahav O., 2004, PASP, 116,345
Connolly A. J., Csabai I., Szalay A. S. et al., 1995, AJ, 110, 2655
Csabai I., Connolly A. J., Szalay A. S, 2000, AJ, 119, 69
Csabai I. et al., 2003, AJ, 125, 580
Cutri R. M., Skrutskie M. F., van Dyk S. et al., 2003, VizieR On-line Data Catalog: II/246
Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339,1195
Hsieh B. C., Yee H. K. C., Lin H. et al., 2005, ApJS, 158,161
Jarrett T. H., Chester T., Cutri R. et al., 2000, AJ, 119, 2498
Koo D. C., 1985, AJ, 90, 418
Li L., Zhang Y., Zhao Y. et al., 2007, Chin. J. Astron. Astrophys. (ChJAA), 7, 448
Loh E. D., Spillar E. J., 1986, ApJ, 303, 154
Nadaraya E. A., 1964, Theory of Probability and its Applications, 9, 141
Gunn S. R., 1998, Support Vector Machines for Classification and Regression, Technical Report, School of Electronics and Computer Science University of Southampton (Southampton, U.K.)
Stoughton C. et al., 2002, AJ, 123, 485
Suchkov A. A., Hanisch R. J., Margon B., 2005, AJ, 130, 2439
Tagliaferri R., Longo G., Andreon S. et al., 2003, Lecture Notes in Computer Science, 2859, 226
Vapnik V. N., 1995, The Nature of Statistical Learning Theory, New York: Springer-Verlag
Vanzella E. et al., 2004, A&A, 423, 761
Wadadekar Y., 2005, PASP, 117, 79
Wang D. et al., 2007, MNRAS, 382, 1601
Wang Y., Bahcall N., Turner E., 1998, AJ, 116, 2081
Watson G. S., 1964, The Indian Journal of Statistics, Series A, 26, 359
Way M. J., Srivastava A. N., 2006, ApJ, 647, 102
Williams S. J., Wozniak P. R., Vestrand W. T. et al., 2004, ApJ, 128, 2965
Xia L. F. et al., 2002, PASP, 114, 1349
York D. G. et al., 2000, AJ, 120, 1579
Zhang Y. X., Cui C. Z., Zhao Y. H., 2002, In: Jean-Luc Starck, Fionn D. Murtagh eds., Astronomical Data Analysis II, Proc. of SPIE, 4847, 371
Zhang Y. X., Zhao Y. H., 2004, A&A, 422, 1113
Zhang Y. X., Zhao Y. H., 2007, Chin. J. Astron. Astrophys. (ChJAA), 7, 289