

Improving Accuracy of Quasars' Photometric Redshift Estimation by Integration of KNN and SVM

Bo Han¹, Hongpeng Ding¹, Yanxia Zhang² and Yongheng Zhao²

¹International School of Software, Wuhan University, Wuhan, P.R.China
email: bhan@whu.edu.cn; 2010282160014@whu.edu.cn

²Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, 100012, Beijing, P.R.China

Abstract. Catastrophic failure is an unsolved problem existing in the most photometric redshift estimation approaches for a long history. In this study, we propose a novel approach by integration of k-nearest-neighbor (KNN) and support vector machine (SVM) methods together. Experiments based on the quasar sample from SDSS show that the fusion approach can significantly mitigate catastrophic failure and improve the accuracy of photometric redshift estimation.

Keywords. Photometric redshifts, K-Nearest-Neighbour, Support Vector Machine, SDSS

Methodology and Results

We analyze the reasons of catastrophic failure of photometric redshift estimation for quasars and point out that the outlier points are resulted by non-linearly separation in Euclidean feature space of input pattern. Therefore we propose a new approach by integration of SVM and KNN methods. By Gaussian kernel function in SVM, we map input pattern from an original Euclidean space into a high dimensional feature space. In this way, many outlier points can be identified by the hyperplane and then corrected. As shown in Figure 1, the experimental results based on quasars from SDSS database indicate that the integration approach can significantly mitigate catastrophic failure and improve the photometric redshift estimation accuracy (RMS of Δz is reduced from 0.258 to 0.246, and the percent when $|\Delta z| < 0.3$ is improved from 87.02% to 90.03%, for $|\Delta z| < 0.2$ from 85.68% to 88.78%, for $|\Delta z| < 0.1$ from 78.70% to 81.70%). While other researchers mitigate catastrophic failure by cross-match of data from multiple surveys, our approach achieves the similar objective only from a single survey and therefore can be applied to much larger data avoiding cross-match efforts.

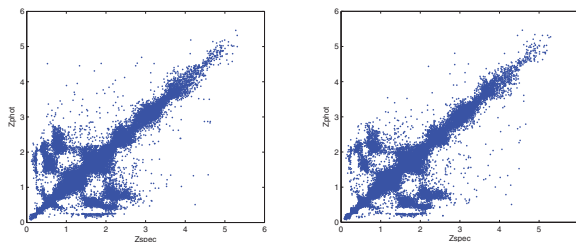


Figure 1. Left: Estimation by KNN; right: Estimation by KNN+SVM