Abstract

According to the standard ΛCDM model, the Universe is composed of dark energy, radiation, dark matter (DM), and baryonic matter. The distribution of total matter (DM + baryons) can be traced through the clustering of large-scale structure (LSS) such as galaxies and quasars. LSS surveys are conducted using both photometric and spectroscopic techniques. Spectroscopic surveys provide precise redshift measurements (spec-zs) from spectra and enable the three-dimensional mapping of LSS. However, they are observationally expensive and time-consuming. In contrast, photometric surveys can observe large areas of the sky, detecting millions of objects to great depths within a relatively short time. Their main limitation is that they do not provide direct redshift measurements. Obtaining spec-zs for all objects in photometric surveys is impractical due to the long exposure times required, which poses a significant challenge for using photometric data in cosmological analyses where redshift is a fundamental quantity.

This thesis addresses this challenge by estimating redshifts from photometric quantities using advanced machine learning algorithms. These estimates, known as photometric redshifts (photo-zs), are less precise than spec-zs. Although photo-zs are not reliable for full three-dimensional clustering due to their uncertainties, they can be effectively used for angular (two-dimensional) clustering measurements. Although redshift information is not needed to measure the angular clustering in observed data, the redshift distribution is essential for modelling this observable. This distribution is derived from photo-zs. Since galaxies are biased tracers of total matter, while dark matter itself is not directly observable, this modelling allows us to estimate the bias and better understand the relationship between visible structures and the total matter distribution in our Universe.

A central contribution of this thesis is the development of the deep learning framework `Hybrid-z` for photo-z estimation. This model combines four-band optical imaging data with nine-band photometric magnitudes through a hybrid architecture of Convolutional Neural Networks and Artificial Neural Networks. Applied to the Kilo-Degree Survey Data Release 4 (KiDS-DR4) bright galaxy catalogue, `Hybrid-z` significantly improves the photo-zs. It achieves a ∼20% reduction in scatter compared to the previous method that relies only on photometric magnitudes rather than imaging data. The model is publicly released as an open-source tool, and the resulting photo-z catalogue for the KiDS-DR4 bright galaxy sample is made available to the community.

`Hybrid-z` is further applied to KiDS-DR4 quasars to construct redshift distributions and analyse their clustering using the angular two-point correlation function. The best-fit scale-independent quasar bias increases from $b \approx 1.6$ at $z \approx 0.6$ to $b \approx 4.0$ at $z \approx 2.2$, following a quadratic trend with redshift. The clustering results suggest that quasars inhabit DM halos with masses $\log_{10}(M_{\rm eff}/h^{-1}M_{\odot}) \sim 12.7$–$12.9$ and peak heights $\nu_{\rm eff}$ rising from ~ 1.5 to 2.9 over the considered redshift range. We also studied the systematic

effects on bias calculation. It shows that stellar contamination has a negligible effect, while the redshift distribution significantly affects the inferred bias, highlighting the need for accurate redshift calibration.

Beyond two-point statistics, this thesis studies the non-Gaussian features of the matter distribution using higher-order statistics derived from the Count-in-Cells method. Measurements of average correlation function and reduced cumulants in photo-z bins of KiDS-DR4 bright sample reveal significant non-Gaussianity and its evolution with redshift. The clustering signal is strongest on small angular scales and decreases with increasing scale, with indications of observational systematics on large scales. Finally, the dependence of clustering on galaxy properties is examined. Galaxies with higher stellar mass and redder colours exhibit stronger clustering, indicating that they preferentially inhabit denser environments.

These results provide insights into the relationship between luminous tracers in KiDS-DR4 and the underlying matter distribution, and demonstrate the potential of using machine learning for photo-z estimation with clustering analysis for next-generation LSS surveys.