Deep Learning Approach to Photometric Redshift Estimation

Krishna Chunduri*[1] and Mithun Mahesh*[1]

[1] *Cambridge Centre for International Research, Ltd*
184 Cambridge Science Park Milton Rd, Milton
*Cambridge CB4 0PZ, United Kingdom**

Reporter: Baisong Zhang (张百松)

2024-06-27

1. Introduction

- Photometric redshift estimation is an essential process in modern astronomy, determining the redshift of celestial objects, such as galaxies and quasars.
- By measuring the object's magnitude in different wavelength filters, such as ultraviolet (u) or green (g), and evaluating the differences in magnitude to determine the object's color (u-g), we can use color values can help estimate redshift for the celestial object*(Newman & Gruen 2022).*

— shedding light on distances for celestial objects

— advancing our grasp on galaxy formation and evolution

1. Introduction

- Traditional methods often employ spectroscopy to determine redshift, utilizing galaxy spectral signature and wavelength shifts.
  - However, this technique can be resource-intensive and expensive.
  - Furthermore, faint celestial objects can pose challenges to spectroscopic observations.
- These drawbacks have led to the emergence of photometric redshift as a viable alternative.
- Photometric redshift estimation harnesses the magnitude of extragalactic objects as observed across multiple filters*(Salvato et al. 2018).*

1. Introduction

- Previous studies have found significant advancements.
- The CANDELS GOODS-S survey, utilizing the HST WFC3 H-band and ACS z-band, has expand our understanding of photometric redshifts*(Dahlen et al. 2013).*
  — the research found a **direct correlation between the source magnitude and the precision of redshift estimation**, emphasizing the role of magnitude in estimation.
- Another approach was utilizing Bayesian methodologies*(Benitez 2000).*
  — By employing prior probabilities and Bayesian marginalization, this method was adept at utilizing previously overlooked data like the expected shape of redshift distributions and galaxy type fractions.
- Given the next generation of surveys from the James Webb Space Telescope (JWST) and Rubin Observatory (LSST), photometric redshift estimation needs a more data-driven approach to accurately predict redshift based on observational data.

1. Introduction

- The primary objective of this paper is to explore novel computational methods that take a data-driven approach to estimation, while increasing accuracy.
- Specifically, this research aims to evaluate the reliability of Fully Connected Neural Networks (FCN) in estimating photometric redshift using magnitude data.
- We aim to create both a **decision tree regression model** and a **FCN** for photometric redshift estimation. Comparison metrics between the two methods will be **RMS values** and overall prediction accuracy.

2. Data

Table 1. Data Set

u	g	r	i	z	redshift
18.27449	17.01069	16.39594	16.0505	15.79158	0.0369225
18.51085	17.42787	16.94735	16.61756	16.46231	0.06583611
18.86066	17.91374	17.56237	17.26353	17.13068	0.1202669
19.38744	18.37505	17.63306	17.25172	17.00577	0.1806593
18.38328	16.59322	15.77696	15.3979	15.08755	0.04035749

- Our study utilized a dataset from the Sloan Digital Sky Survey*(Kollmeier et al. 2017)* with 50,000 celestial objects.
- The 5 bands - u, g, r, i and z - represent different wavelengths of light from each galaxy or quasar*(Wyder et al. 2007).* Alongside the magnitudes, the dataset came with redshift value labels for each object.
- These redshifts were obtained from spectroscopic measurements from SDSS.
- The first 5 rows can be found in Table 1.

2. Data

- In Fig. 1, the distribution of the redshift and magnitude values is illustrated.
  - For preprocessing, we performed sigma-clipping using a sigma value of 3 standard deviations to remove outliers while retaining 95% of the data.
  - Additionally, we removed redshift values less than zero as these are not physical.
  - As a result, we ended with a dataset of 47,484 celestial objects out of the original 50,000.

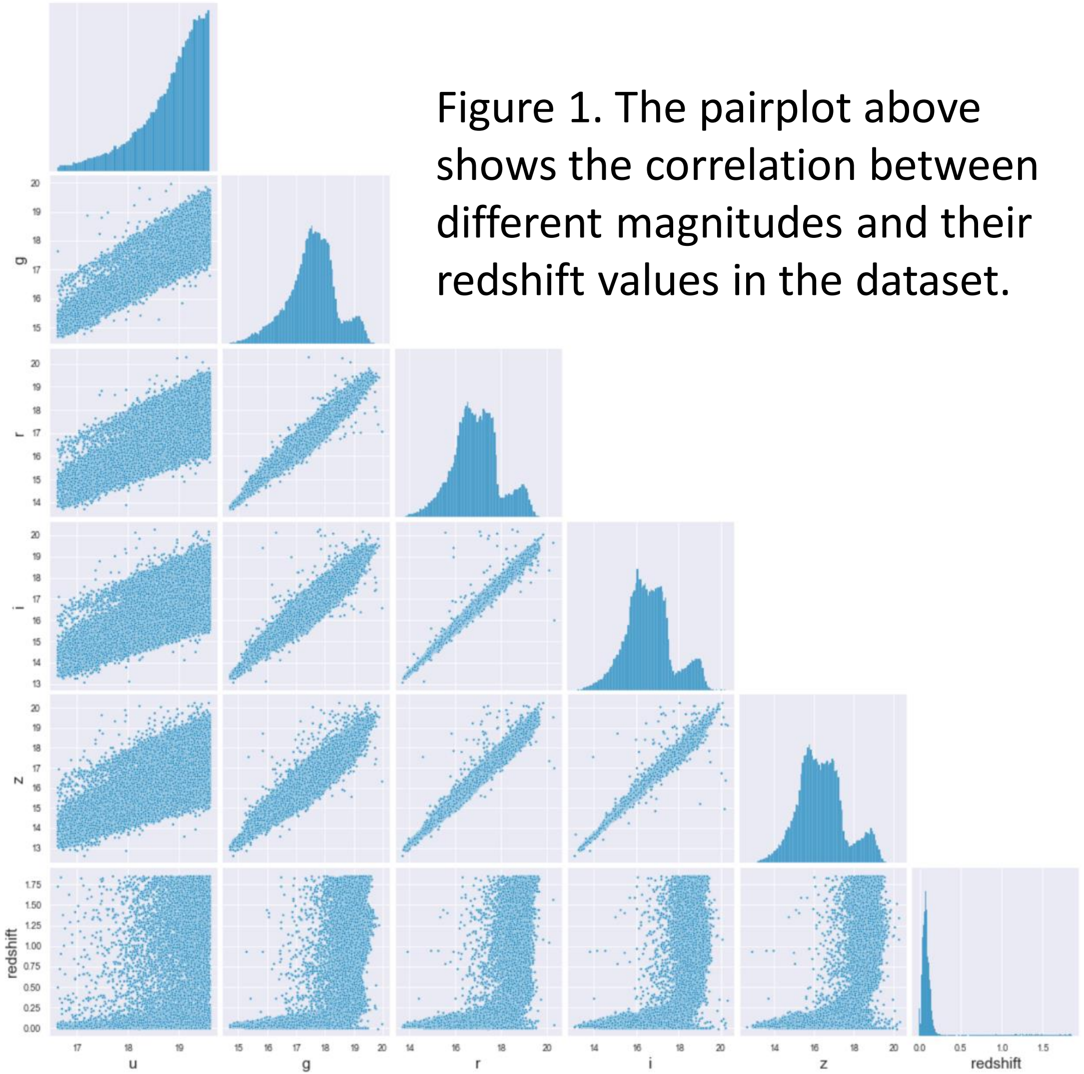


Figure 1. The pairplot above shows the correlation between different magnitudes and their redshift values in the dataset.

3. Methodology

- Decision tree regressor.
  - The decision tree regressor works by partitioning the datasets into small subsets. Each split is based on the value of the input features.
  - Our features consisted of the 5 bandpass filters (u, g, r, i, z) as well as the colors formed by their magnitude differences ($u - g$, $g - r$, $r - i$, $i - z$).
  - After splitting the data, we arrive at leaf nodes where the redshift values are as similar as possible.
  - Each leaf of the tree then predicts the average redshift of the instances that fall into it.
  - The model is simple and transparent, but doesn’t produce very good results in terms of RMS（0.16） and prediction.

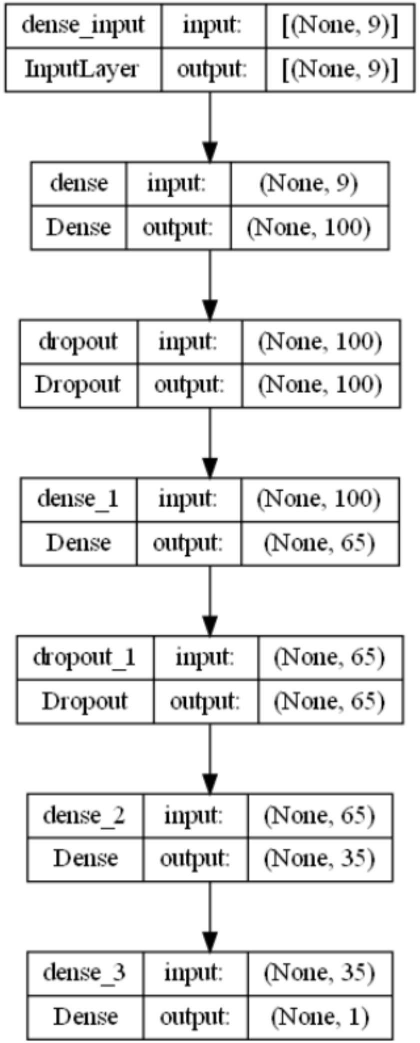
3. Methodology

- We chose a fully connected neural network that used the Adaptive Moment Estimation Optimizer[(Kingma & Ba 2017)] in order to create a regression model to predict redshift.

$$Input = [m_u, m_g, m_r, m_i, m_u - m_g, m_g - m_r, m_r - m_i, m_i - m_z].$$

  — We used the ReLU activation function[(Agarap 2019)] which worked better than the sigmoid function to account for redshift predictions with values greater than 1 as well as to improve efficiency of the network.

  — Lastly we added a dropout rate of 0.2 to prevent overfitting after each layer.

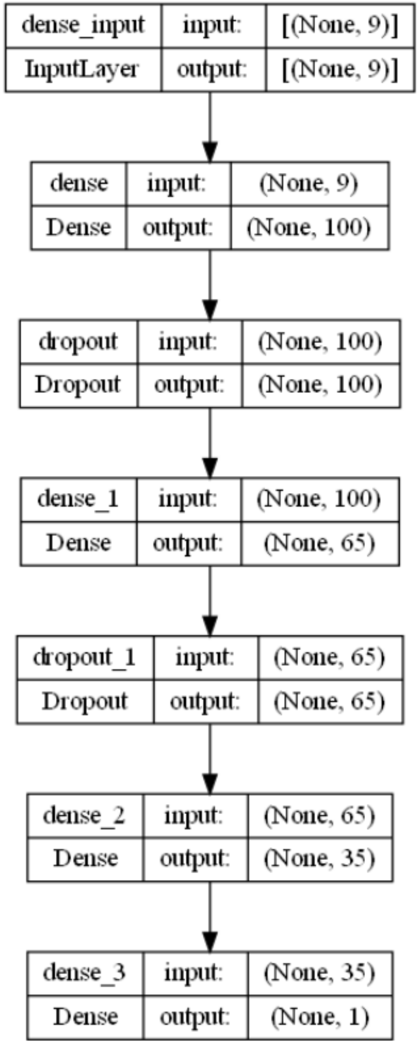


3. Methodology

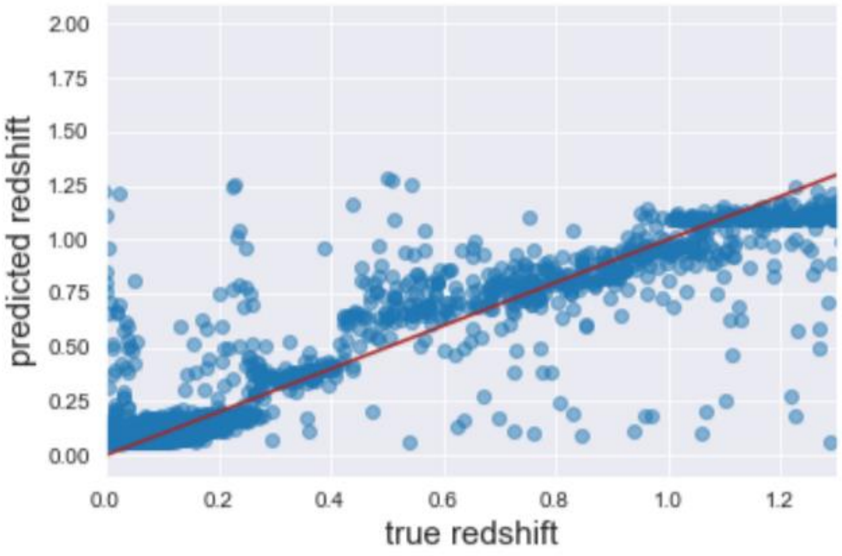
- We minimise the mean squared error (MSE) as the loss function in our neural network

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y})^2,$$

— where y is the true redshift, ˆy is the predicted redshift, and n is the number of objects in a batch of the training set.



4. Result



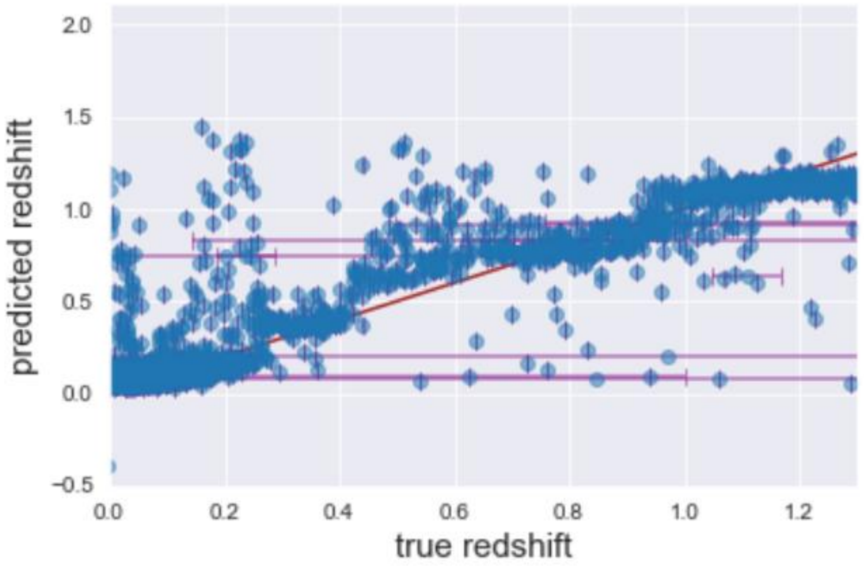


Figure 3. The chart above shows true redshift vs predicted redshift correlation. There are few outliers, with majority of predictions being close to the best fit line.

Figure 4. The chart above shows error bars along with the true redshift vs predicted redshift graph.

4. Result

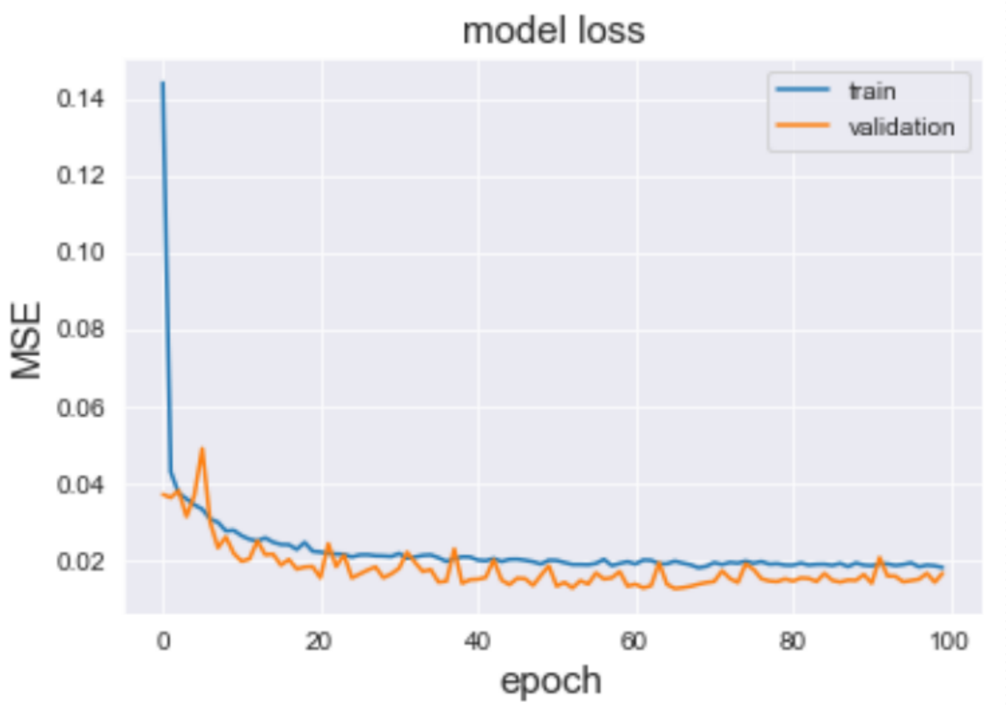


Figure 5. The chart above shows the learning curve. The loss of the training function followed the same trajectory as that of the validation set, stabilizing and reaching an equilibrium, indicating a good fit.

5. Discussion

- The empirical evidence from our study has not only demonstrated a data-driven approach but has shed light on incredibly efficient methods for photometric redshift estimation.
- Compared to the traditional decision tree regression models, the Fully Connected Neural Network showcases a clear edge in estimating redshifts.
- Moreover, as the acquisition of SED templates becomes increasingly challenging, the need for approaches that can efficiently utilize raw astronomical data will become more pressing.

6. Conclusions

- Our study underscores the untapped potential of data-driven methodologies in photometric redshift estimation, particularly highlighting the superior capabilities of FCNs over decision tree regressors.
- While traditional methods, such as decision tree regression, continue to hold value, the evolving landscape of computational methods offers new opportunities for precision and discovery in our universe.
- As we anticipate the demands of next-generation astronomical surveys, including those from the James Webb Space Telescope and Large Synoptic Survey Telescope*(Ivezi ́c et al. 2019)*, these data-centric approaches will be pivotal in unveiling the redshift to find distances for far-away celestial objects, from quasars to galaxies.