Refined Redshift Regression in Cosmology with Graph Convolution Networks

Robert Beck
Institute for Astronomy
University of Hawai'i at Mānoa
Honolulu, HI 96822
beckrob@ifa.hawaii.edu

Yannik Glaser
Information and Computer Sciences
University of Hawai'i at Mānoa
Honolulu, HI 96822
yglaser@hawaii.edu

Peter Sadowski
Information and Computer Sciences
University of Hawai'i at Mānoa
Honolulu, HI 96822
peter.sadowski@hawaii.edu

Istvan Szapudi
Institute for Astronomy
University of Hawai'i at Mānoa
Honolulu, HI 96822
szapudi@ifa.hawaii.edu

Abstract

The redshift of a galaxy uniquely determines its distance in our expanding universe. The colors of a galaxy measured in different bands constrain the approximate "photometric" redshift, although the prediction models are imprecise and improving these estimates is a hot topic in cosmology. We use machine learning models to refine the photometric redshifts of galaxies using their spatial structure organized into clusters, filaments, and walls. In particular, we test the hypothesis that additional information from the "neighborhood" of a galaxy sharpens our photometric redshift estimate. We demonstrate that a graph convolutional neural network — trained on a data set of high-resolution redshift observations on a small region of the sky — captures this information by learning to predict the redshift of all the galaxies in a viewing region simultaneously, improving the performance over single-galaxy redshift prediction by 10% median absolute deviation on a held-out region of the sky.

1 Introduction

The redshift z of a galaxy corresponds to the fractional change of wavelength due to the cosmological expansion rate of the universe. Since this expansion rate is a unique function of the distance in a particular cosmological model, measuring redshifts are virtually equivalent to measuring radial distances. Together with the direction of the galaxy, the redshift measurements allow us to create three-dimensional maps of the distribution of galaxies. The cosmological information (the number of different ingredients, the curvature, and the age of the universe) of such three-dimensional data is about five times more than a corresponding two-dimensional map using only the directions of galaxies.

The redshift of an astronomical object can generally be determined very accurately by identifying absorption or emission lines of certain elements in high-resolution spectra. However, spectroscopy is extremely expensive. Broad-band photometric observations, on the other hand, are far less time-consuming in terms of telescope time, and measure broad wavelength ranges from a light source to create a very rough representation of a spectrum. For this reason, spectroscopic catalogs are often utilized as labelled training data sets to build statistical models that reconstruct the redshift of a galaxy from a small number of colors — typically five or more. This is challenging due to measurement

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

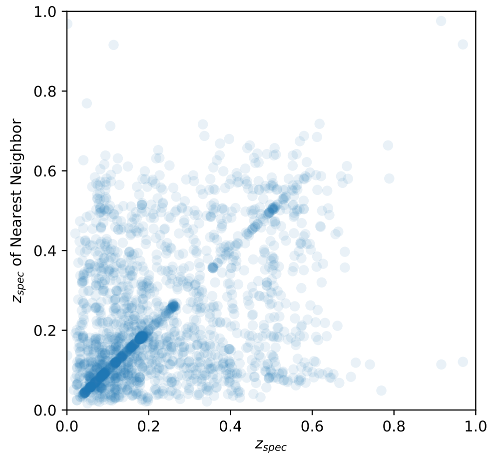

Figure 1: Spectroscopic redshift (z_{spec}; unitless) of 2000 galaxies vs. the spectroscopic redshift for the nearest galaxy, in terms of Haversine distance in the viewing field. The line of points along the diagonal shows the high correlation of galaxies that are in the same cluster. Photometric redshift estimates are much less accurate than spectroscopic measurements, so the goal of this work is to refine photometric redshift estimates using the correlations between neighboring galaxies. The challenge lies in identifying *which* of a galaxy's neighbors are in the same cluster, given only noisy photometry data.

errors, modeling errors, and degeneracies, as different redshift can correspond to similar areas in a five-dimensional color space, but improving these redshift estimates is ripe for rewards: the amount of cosmological information increases almost linearly with the inverse of the photometric redshift errors. The reason for this is that much of the information is contained in features (BAO: baryonic acoustic oscillations) on scales that are comparable to the errors of photometric redshifts determined by state of the art methods, which include machine learning techniques such as random forest, support vector machines, and deep learning methods.

Since galaxies form complex structures, clusters, filaments, and walls, their redshifts are not independent of each other. Galaxies in the same cluster have essentially the same redshift (Figure 1), with the cluster size two orders of magnitude smaller than present photometric redshift errors. Thus the primary challenge is to identify which of a galaxy's neighbors are in the same cluster from both high-variance photometery redshift measurements in the line of sight dimension and precise measurements in the orthogonal dimensions. Filaments and walls constrain the photometric redshift on a one and two-dimensional locus, respectively.

There have been previous attempts to use the spatial distribution of galaxies to estimate empirical redshift distributions using galaxy clustering [1]. Indeed, it has been shown that if a subset of galaxies has direct redshift measurements, they can improve the photometric redshift estimate of nearby galaxies [2]. However, the interdependence of photometric redshifts is complex and not fully captured by existing methods. In this paper, we show that graph convolutional networks provide an elegant method of modeling these dependencies, and increase the accuracy of photometric redshift estimates over the state of the art methods based on individual colors.

2 Methods

2.1 Graph Neural Network Architectures

The photometric galaxy observations under discussion are essentially very large, sparse, multicolor images. Due to this sparsity, standard 2D convolutions used for computer vision tasks would be computationally expensive even at greatly-reduced resolution, and furthermore would *not* be invariant to rotations of the image. In contrast, *graph* convolution networks provide an efficient method for modeling the relevant spatial structure of the data while also maintaining a sparse representation.

By representing each galaxy as a node in a graph with edges denoted by the pair-wise Haversine distance — the geodesic distance on a sphere — we achieve a much more compact representation of the data and convolution operators that are invariant to both translations *and* rotations of the input. This approach has proven useful for high-dimensional detectors in particle physics [3, 4].

Many formulations of graph convolution networks exist [5, 6], but the key idea is to associate some representation with every node in a graph and apply a (learned) function to every node. This function maps the representation of a node, those of its neighbors, and the edge annotations to a new representation. In this work we tested three different graph convolution architectural motifs for the task of incorporating galaxy cluster information into photometric redshift estimation.

1. **Graph Convolution with Attention**: A weighted average is computed over the learned representations of a galaxy's neighbors, where the weighting is determined by a dot product of the galaxy's own representation with that of its neighbors.
2. **MoNet-style Architecture**: This model generalizes Attention by replacing the dot-product with a parameterized, pair-wise, weight function [6]. This is more computationally expensive than standard attention, but we found it to be an effective architecture for this application. Our architecture does not follow the MoNet architecture from the paper exactly — instead of using a Gaussian weight function, we use a small neural network — so we refer to this as a "MoNet-style" architecture.
3. **Recurrent Message Passing**: In the present application, each node encodes the redshift estimate of a galaxy, and graph convolution layers help to refine these estimates by incorporating information ("messages") about neighboring galaxies. As suggested in [7, 8], a *recurrent* architecture can apply the same function repeatedly — iteratively updating each node until convergence.

2.2 Data Set

To date, the Sloan Digital Sky Survey [SDSS; 9, 10] has provided the most extensive collection of spectroscopic observations of galaxies [11, 12], as it includes close to 2 million objects over an impressive area of $\approx 14,000$ square degrees on the sky, extending out to a redshift of $z \approx 0.6$. The combined characteristics of depth, area, and source count of the SDSS spectroscopic dataset make it an ideal data set for training and evaluating spatial photo-z methods.

We use the data set described by [1], in which each galaxy is described by five photometric features: the r-band magnitude, and $u-g$, $g-r$, $r-i$ and $i-z$ colors. These were scaled to zero mean and unit standard deviation. Spatial information is encoded as the right ascension (RA) and declination (Dec) coordinates of each galaxy.

Because the southern sky is less homogeneous, we limit our spatial analysis to the northern sky, specifically the range $\mathrm{RA} \in [130°, 230°]$ and $\mathrm{Dec} \in [5°, 55°]$, avoiding the boundaries of the survey. This sky patch was subdivided into a training ($\mathrm{RA} \in [130°, 183°]$) and validation ($\mathrm{RA} \in [187°, 230°]$) data set, ensuring no overlap of either single galaxies or clusters. This range is large enough that local large-scale structures are unlikely to create systematic differences in the two data sets.

Training is performed using mini-batch stochastic gradient, where each mini-batch consists of a small region of the sky and all pair-wise distances are included in the graph. The training region of the sky is divided into tiles of $1.6° \times 1.6°$, then each epoch of training consists of iterating through the square tiles (in random order); a point is sampled from a uniform distribution over this tile, and all galaxies within a radius of $0.9°$ are included in the mini-batch. As a regularization method, we also explored randomly dropping out $0-10\%$ of the objects from each circular mini-batch. During validation, we again divide the sky patch into $1.6° \times 1.6°$ tiles, but the circular cutouts are always placed on the tile center points, and no objects are dropped. Thus, validation is always performed on the same set of mini-batches in the same data format as the training data. The primary performance metric is the median absolute deviation (MAD), where the median is taken over the validation set.

$$MAD = \mathrm{median}\left(\frac{z_{phot} - z_{spec}}{1.0 + z_{spec}}\right) \tag{1}$$

3 Results

In experiments, we found that it was difficult for a standard attention model to efficiently incorporate neighborhood information. We reasoned that this was due to the need for selecting the few neighboring galaxies that are in the same cluster and assigning them high weight, while ignoring the many distractor galaxies. Thus, we focused on MoNet-style architectures, which are more flexible in how they weight neighbors.

The best architecture was constructed in three phases. First, a single-object model was trained that did not use any neighborhood information. This model consisted of the five photometry inputs, three hidden layers of 512 ReLU units with 10% dropout in the last layer, and a single softplus output trained with mean squared error. This achieved a MAD score of 0.0125 on the validation set, and served as a baseline for comparing performance with more complex models. Next, a MoNet-style architecture was trained to aggregate neighborhood information, using the predictions of the first network as inputs. The networks learns a function that computes pair-wise weights from (1) photometry features of two galaxies, (2) their estimated z_{phot} from the single-object network, and (3) the Haversine distance between the galaxies. The weight function has three hidden layers of 256 ReLU units and 10 linear outputs. These 10 weights are used to compute weighted averages over a galaxy's neighbors' z_{phot}, and a final segment of the network aggregates these ten weighted averages with three more hidden layers of 256 ReLU units followed by a softplus output. This model was trained while fixing the weights of the single-object subnetwork. In the third and final phase, the weights of the single-object subnetwork were released and the entire model was fine-tuned together, achieving a validation set MAD of 0.0112 — a 10% improvement over the single-object network (Figure 2).

Additional experiments were carried out with recurrent versions of architecture above and tuned using the SHERPA hyper-parameter optimization framework [13], albeit with smaller layers so that training could be performed on a single GPU. So far, we have been unable to improve performance using this approach, but we expect this to be a promising direction of future research.

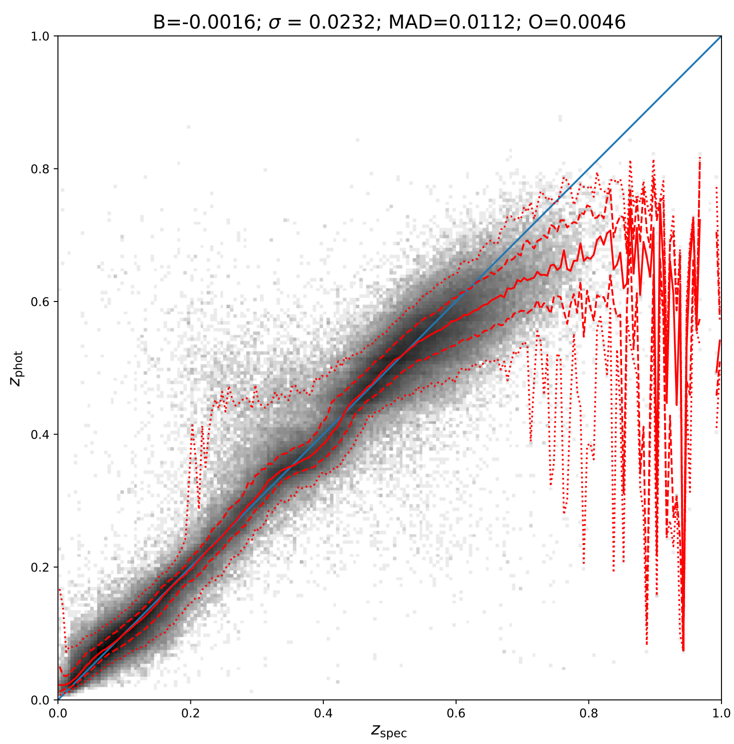

Figure 2: Predicted redshift from photometric observations using the MoNet-style model (z_{phot}) vs. ground truth measurement via spectroscopy (z_{spec}) of the validation set. For each z_{spec} bin, we also plot the median (solid line), the 68% confidence interval (dashed lines), and 95% confidence interval (dotted lines) of the empirical z_{phot}.

4 Discussion and Conclusion

Estimating redshift is critical for mapping out the 3D structure of the universe and providing the data used to fit cosmological models. This work demonstrates the ability of graph convolutional neural networks to increase the precision of photometric redshift estimates by incorporating spatial information from neighboring galaxies. In our experiments with different graph convolution architectures, we found that a MoNet-style architecture was able to reduce the median absolute deviation (MAD) from 0.0125 — for a neural network trained on only single-object photometry features — to 0.0112, a 10% decrease.

However, we believe the neighborhood information captured in these preliminary experiments is incomplete, and that there is an opportunity for more sophisticated methods. A promising approach is to apply recurrent architectures for iterative refinement, and we are currently exploring recurrent versions of the best architecture.

References

[1] R. Beck, L. Dobos, T. Budavári, A. S. Szalay, and I. Csabai. Photometric redshifts for the SDSS Data Release 12. , 460:1371–1381, August 2016.

[2] M. A. Aragon-Calvo, Rien van de Weygaert, Bernard J. T. Jones, and Bahram Mobasher. Submegaparsec individual photometric redshift estimation from cosmic web constraints. , 454(1):463–477, Nov 2015.

[3] Isaac Henrion, Johann Brehmer, Joan Bruna, et al. Neural message passing for jet physics. 2017.

[4] Nicholas Choma, Federico Monti, Lisa Gerhardt, et al. Graph neural networks for icecube signal classification. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 386–391. IEEE, 2018.

[5] Marco Gori, Gabriele Monfardini, and Franco Scarselli. A new model for learning in graph domains. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 2, pages 729–734. IEEE, 2005.

[6] Federico Monti, Davide Boscaini, Jonathan Masci, et al. Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5115–5124, 2017.

[7] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2008.

[8] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1263–1272. JMLR. org, 2017.

[9] D. G. York, J. Adelman, J. E. Anderson, Jr., et al. The Sloan Digital Sky Survey: Technical Summary. , 120:1579–1587, September 2000.

[10] D. J. Eisenstein, D. H. Weinberg, E. Agol, et al. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. , 142:72, September 2011.

[11] M. A. Strauss, D. H. Weinberg, R. H. Lupton, et al. Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample. , 124:1810–1824, September 2002.

[12] K. S. Dawson, D. J. Schlegel, C. P. Ahn, et al. The Baryon Oscillation Spectroscopic Survey of SDSS-III. , 145:10, January 2013.

[13] Lars Hertel, Julian Collado, Peter Sadowski, and Pierre Baldi. Sherpa: Hyperparameter optimization for machine learning models. 2018.