Composite Spectral Energy Distribution and Clustering Methods

Ming-Feng Ho[1]

[1]Department of Physics and Astronomy, University of California, Riverside; email: mho026@ucr.edu

Xxxx. Xxx. Xxx. Xxx. YYYY. AA:1–6

https://doi.org/10.1146/((please add article doi))

Copyright © YYYY by Annual Reviews. All rights reserved

Keywords

galaxies: evolution, methods: data analysis, methods: Bayesian non-parametric

Abstract

Composite SEDs and Bayesian non-parametric clustering. Photometry and medium bands: surveys Spectral Energy Distributions: fitting template, FAST, EAZY Composite SEDs: evolution from grouping methods Bayesian non-parametric on functional data: 1. Dirichlet Processes for clustering 2. Gaussian Processes on Spectral data 3. Clustering on functional data

Contents

1. INTRODUCTION

improvement of Photometry data on redshift rage 3-4.

2. GALAXY EVOLUTION IN TERMS OF COMPOSITE SPECTRAL ENERGY DISTRIBUTIONS (SEDs)

2.1. Medium-Band Photometry

This is dummy text.

2.2. Fitting Template of Spectral Energy Distributions (SEDs)

This is dummy text. This is dummy text. This is dummy text. This is dummy text.

2.3. Composite Spectral Energy Distributions (SEDs)

3. BAYESIAN MACHINE LEARNING FOR CLUSTERING AND MODELING SPECTRAL DATA

Bayesian machine learning is a branch of machine learning which aims to solve machine learning problems in a Bayesian perspective. Instead of optimizing the parameters of interest from data using an empirical loss function (e.g., a least-squared function), Bayesian methods build generative models to randomly sample data from parameters and try to maximize the likelihood between observed data and hidden parameters (Barber 2012).

The difference between Bayesian statistics and Bayesian "machine learning" is that Bayesian "machine learning" is trying to approximate *non-linear* functions (Bishop & Tipping 2003). After the publish of Rasmussen & Williams (2005), learning unknown complicated functions from observed data using *Gaussian processes* (GP) became popular.

3.1. Modeling Spectral Data using Gaussian Processes (gp)

A *Gaussian process* is a bunch of random variables, and any finite subset of these random variables is a joint Gaussian distribution (Rasmussen & Williams 2005). GP could be a powerful tool to model any kind of functional data (continuous data) in a non-parametric way. By non-parametric, it actually means we use infinite many parameters to describe our function (Gelman et al. 2014). GP could be treated as a random function (or a stochastic process) which draws samples from the n-dimensional distribution,

$$\mu(x_1), ..., \mu(x_n) \sim Normal((m(x_1), ..., m(x_n)), K(x_1, ..., x_n)). \quad 1.$$

The construction of a GP could be considered as finding the mean function ($m(\vec{x})$) and a suitable covariance function ($K(\vec{x}, \vec{x}')$). In normal cases, a zero mean is usually used as a prior for GP regressions. For $K(\vec{x}, \vec{x}')$, pre-defined covariance functions (e.g., squared potential function $\exp(\frac{-r^2}{2\ell^2})$) are often been implemented. However, the usage of GP in modeling functional data will also be restricted by the intrinsic properties of the covariance functions. Learning a suitable covariance function is the most crucial part of machine learning in GP.

Finding a suitable choice of covariance often reflects our interpretations of the characteristics of our data (Rasmussen & Williams 2005). For example, the usage of the squared potential function $\exp(\frac{-r^2}{2\ell^2})$ implies the assumption that we believe each point on the function would have less impact to each other if they are far away on the functional space. Therefore, we need a special kind of covariance function to suit our purpose of modeling spctral data.

Garnett et al. (2017) took a machine learning approach to learn the covariance function, with a wavelength range from Ly_∞ to $\mathrm{Ly}\alpha$, from training data (quasar spectra). The optimization choice was to firstly decompose covariance matrix with (Garnett, Ho & Schneider 2015),

$$\mathbf{K} = \mathbf{M}\mathbf{M}^T, \quad 2.$$

and then use the first 10 principle components of the flux of quasar spectra, $\mathbf{Y}$, to constitute the matrix $\mathbf{M}$. The optimization was done by maximizing the log likelihood, $\mathcal{L}$, of the data by given $\mathbf{M}$ and absorption noise ω,

$$\mathcal{L}(\mathbf{M}, \omega) = \log p(\mathbf{Y} \mid \lambda, \mu, \mathbf{M}, \mathbf{N}, \omega, z_{\mathrm{qso}}, Model). \quad 3.$$

The goal of optimizing above function is to find optimal covariance matrix, $\mathbf{M}$, and absorption parameter, ω, with some given conditions. Those conditions are: a given mean vector μ, the noise on the spectra $\mathbf{N}$, the redshift of the QSO, and with a given model. In a perspective of generative modeling, optimizing the data likelihood implies we are trying to find a covariance matrix to better generate our spectral data.

The covariance matrix built in Garnett et al. (2017) with a wavelength range from Ly_∞ to $\mathrm{Ly}\alpha$ is in Fig 1. The scale in Fig1 represents the strength of correlations between pairs of rest-frame wavelengths on the QSO spectra. The features of Lyman series are distinct. The off-diagonal term demonstrates the correlations of pairs of corresponding emission lines.

The mean function of GP modeling in Garnett et al. (2017) is simply stacking the spectra of training data,

$$\mu_j = \mathrm{median}(y_{ij}), \quad 4.$$

where y_{ij} are the fluxes for spectrum. Fig 2 shows the mean function with a range from Ly_∞ to $\mathrm{Ly}\alpha$. The features emission line of Lyman series are also visible in the figure.

Generally, the GP model for spectral data could be described as

$$p(\mathbf{y} \mid \lambda, \mathbf{v}, \omega, z, Model) = Normal(\mathbf{y}; \mu, \mathbf{K} + \mathbf{\Omega} + \mathbf{V}), \quad 5.$$

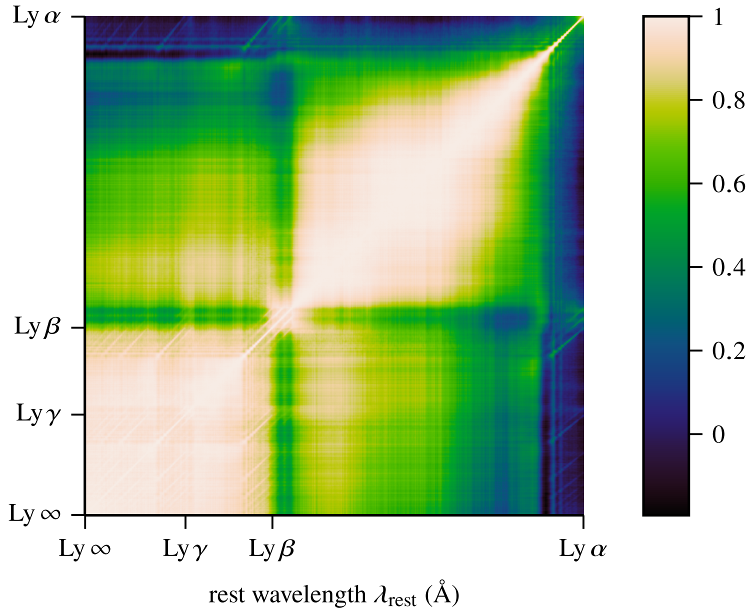

Figure 1

Covariance function for quasar spectra in Garnett et al. (2017)

where $\mathbf{y}$ is the observed flux of the spectrum, λ is the spectroscopic grids we chose to bin the flux, $\mathbf{v}$ is the instrumental noise given by the observed data (it is SDSS QSO catalogue (Pâris, I. et al. 2012) in Garnett et al. (2017)), z is the redshift dependence of the GP model, and ω is the absorption redshift dependence (it was used to model Lyα forest in Garnett et al. (2017)).

The beauty of generative modeling the spectrum using GP framework is that we are able to fully control the modeling of instrumental noise and redshift dependence uncertainties. In addition, the whole framework is transparent and flexible, which implies it is interpretable and future improvements are achievable.

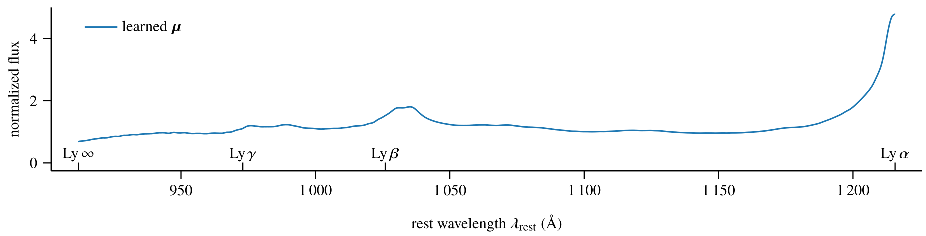

Figure 2

Mean function for quasar spectra in Garnett et al. (2017)

SUMMARY POINTS

1. *Gaussian Processes.* A flexible Bayesian non-parametric framework which allows us to model any kind of function.
2. Learned covariance matrix **K**. To model spectral data, we can use the covariance function via optimizing the covariance function using training data.

3.2. Possibility: Dirichlet Processes combined with Gaussian Process for Modeling Composite SEDs

SUMMARY POINTS

1. Summary point 1. These should be full sentences.

FUTURE ISSUES

1. Future issue 1. These should be full sentences.

ACKNOWLEDGMENTS

Acknowledgements, general annotations, funding.

LITERATURE CITED

Barber D. 2012. *Bayesian Reasoning and Machine Learning*. New York, NY, USA: Cambridge University Press

Bishop C, Tipping ME. 2003. Bayesian regression and classification. Advances in Learning Theory: Methods, Models and Applications

Garnett R, Ho S, Bird S, Schneider J. 2017. *Monthly Notices of the Royal Astronomical Society* 472:1850–1865

Garnett R, Ho S, Schneider J. 2015. *Finding Galaxies in the Shadows of Quasars with Gaussian Processes*. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15. JMLR.org

Gelman A, Carlin JB, Stern HS, Rubin DB. 2014. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd ed.

Pâris, I., Petitjean, P., Aubourg, É., Bailey, S., Ross, N. P., et al. 2012. *A&A* 548:A66

Rasmussen CE, Williams CKI. 2005. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press