Quantifying correlations between galaxy emission lines and stellar continua

R. Beck, L. Dobos, C.W. Yip, A.S. Szalay and I. Csabai
2016
astro-ph: 1601.0241
• Introduction / Technique
• Data
• Emission line fitting
• Reconstruction of emission lines
• BPT diagram
• Stochastic Recipe for emission lines
• Conclusions
Introduction
Introduction

- In general, stellar population synthesis models are very successful in explaining the galaxy SED (e.g. Bruzual & Charlot 03, Maraston+12, Vazdekis+12).
- However, they do not take into account the emission lines.
- Emission lines have their origin from the excited ISM gas and/or AGNs.
- As consequence, pure stellar population models cannot account for observations in broad-bands filters.
Introduction

- To explain a galaxy spectrum, it is necessary to couple the stellar population synthesis and the physics of emission lines.

How to predict the emission lines of a certain galaxy?
Introduction

- To explain a galaxy spectrum, it is necessary to couple the stellar population synthesis and the physics of emission lines.

**How to predict the emission lines of a certain galaxy?**

**Solution 1** - Modeling everything, it is necessary a detailed hydrodynamic model (shock-heating of the ISM, dust emission, etc). (e.g., CLOUD, Le Phare codes)

**Solution 2** - Based on some observations and statistical techniques, one can **empirically estimate** the emission lines based on the stellar continuum (basically stars).
Introduction

- To explain a galaxy spectrum, it is necessary to couple the stellar population synthesis and the physics of emission lines.

How to predict the emission lines of a certain galaxy?

Solution 1
- Modeling everything, it is necessary a detailed hydrodynamic model (shock-heating of the ISM, dust emission, etc). (e.g., CLOUD, Le Phare codes)

Solution 2
- Based on some observations and statistical techniques, one can empirically estimate the emission lines based on the stellar continuum (basically stars).

Once the emission lines are predicted, is it possible to recover the BPT diagram classification?
Introduction

● Ok, empirical solution...but how?

● Part 1- to characterize the stellar continuum: Principal Component Analysis (PCA)

● Eigenspectra are primarily correlated with the emission line strengths and secondly with the continuum slope.

● In order to characterize the continuum, the emission lines have to be modeled/subtracted with minimal residuals.
Introduction

- Ok, empirical solution...but how?

- Part 2- Predict equivalent widths (EWs) based on the stellar continua (PCA projection) using the Machine Learning (ML) approach.
  - Local linear regression
  - K-means clustering
  - Support vector machine (SVM)

- The validity of the results is limited to the training set's coverage of the parameters space (luminosity, redshift, metallicity, etc).
Introduction

- Ok, empirical solution...but how?

Measuring EWs

Projected continua

Machine Learning

Predicted EWs

http://www.astroml.org/sklearn_tutorial/dimensionality_reduction.html
Data
Data

- **Galaxy Sample from the SDSS/DR7**
  - S/N ratio higher than 5 \((r \sim < 19)\).
  - 11 emission lines are measured
- 220 deg. < ra < 230 deg.
- This selection prefers young, small and faint, low-metallicity galaxies at low redshift.
- Final sample: **13788 galaxies**

<table>
<thead>
<tr>
<th>Line</th>
<th>(\lambda_{\text{vac}}) (Å)</th>
<th>Line</th>
<th>(\lambda_{\text{vac}}) (Å)</th>
<th>Line</th>
<th>(\lambda_{\text{vac}}) (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O(\text{III})</td>
<td>3727.09</td>
<td>H(\gamma)</td>
<td>4341.68</td>
<td>O(\text{i})</td>
<td>6365.54</td>
</tr>
<tr>
<td>O(\text{II})</td>
<td>3729.88</td>
<td>O(\text{III})</td>
<td>4364.44</td>
<td>N(\text{II})</td>
<td>6529.03</td>
</tr>
<tr>
<td>H(\beta)</td>
<td>3798.98</td>
<td>H(\beta)</td>
<td>4862.68</td>
<td>N(\text{II})</td>
<td>6549.86</td>
</tr>
<tr>
<td>H(\alpha)</td>
<td>3836.47</td>
<td>O(\text{III})</td>
<td>4932.60</td>
<td>H(\alpha)</td>
<td>6564.61</td>
</tr>
<tr>
<td>H(\epsilon)</td>
<td>3890.16</td>
<td>O(\text{III})</td>
<td>4960.30</td>
<td>N(\text{II})</td>
<td>6585.27</td>
</tr>
<tr>
<td>H(\delta)</td>
<td>3971.20</td>
<td>O(\text{III})</td>
<td>5008.24</td>
<td>S(\text{II})</td>
<td>6718.29</td>
</tr>
<tr>
<td>S(\text{II})</td>
<td>4072.30</td>
<td>H(\alpha)</td>
<td>5877.65</td>
<td>S(\text{I})</td>
<td>6732.67</td>
</tr>
<tr>
<td>H(\delta)</td>
<td>4102.89</td>
<td>O(\text{I})</td>
<td>6302.05</td>
<td>(_)</td>
<td>(_)</td>
</tr>
</tbody>
</table>

Table 1. List of the fitted nebular emission lines.
Emission line fitting
Emission line fitting

- Method of fitting continua - non-negative linear combination of templates.

- Fitting with BC03 + extinction law + intrinsic velocity dispersion

- Known discrepancies between the continuum models and SDSS spectra (it originates from the imperfect models)

- High-pass filter -> 50A wide rolling median filter -> remove incorrect background subtraction.

- **Noise-limited fitting**: Compare rms with background rms, 2x larger then go to more complex model.

\[
F(\lambda) = A \cdot e^{-\frac{(\lambda-\lambda_0)^2}{\sigma^2}} \quad \rightarrow \quad F(\lambda) = A \cdot e^{-\frac{(\lambda-\lambda_0)^2}{\sigma_a^2}} + B \cdot e^{-\frac{(\lambda-\lambda_0)^2}{\sigma_b^2}} \quad \rightarrow \quad F(\lambda) = A \cdot e^{-\frac{(\lambda-\lambda_0)^2}{\sigma_a^2}} + B \cdot e^{-\frac{(\lambda-\lambda_0)^2}{\sigma_b^2}}
\]

1. Necessary for broad wings
2. Assymetric line
Emission line fitting

Template fitting
(10 BC03 templates)

Template fitting
(10 BC03 templates) +
50A low-pass filter
Emission line fitting

Comparison with other works (Brinchmann+04, line ratios imposed)

weak lines are slightly higher
PCA projection

- PCA projection of the stellar continuum from the fitted model spectra instead of the observed one.
- PCA was done in the wavelength range of 3722Å - 6761Å.
- 1st eigenvector is sensitive of the galaxy colour.
- 2nd and 3rd eigenvectors are quite similar, but the latter is more prominent of absorption lines.
- 4th eigenvector - > width of absorption lines.
Reconstruction of emission lines

- The goal is to empirically estimate the EWs from a continuum principal components (first 5 PCs).

- Local linear regression with k-NN: (k=30)

- Emission lines are much better reconstructed in SF galaxies (strong connection between stellar pop. and ISM).

- \([\text{[OIII]}]\) -> this AGN indicator has a problematic reconstruction.

\[ r_i \approx c_i + a_i d_i. \]  
\[ \chi^2 = \sum_{j \in NN} \frac{(r_j - c_i - a_id_j)^2}{w_j}, \]
Reconstruction of emission lines

- **SDSS magnitudes** are contaminated with emission lines which might result in some correlation with EWs.

- Instead of spectrum, the emission lines can be reconstructed from SDSS mags (no k-correction).

- Local fitting + kNN is necessary to reconstruct the BPT diagram from either continuum principal components or broad-band magnitudes.

- How about the BPT diagram?
BPT diagram

- (a) PCA continuum + Local linear regression with kNN (k=30).
- (b) SDSS magnitudes.
- (c) PCA + kNN (k=30 randomly selected galaxies).
- (d) Local linear regression + shuffled PCs.
BPT diagram
Revisiting Star-forming / AGN separation

- SF and AGNs cannot be clearly separated as a bimodal classification.

- Support Vector Machine (SVM)

- Select galaxies with high confidence of SF or AGNs.

- They used 5 continuum PCs + 4 PCs of EWs of the training set.

- SVM reproduced the empirical segregation of K+03. Only 6% is scattered to the opposite region.
They propose a simple stochastic recipe to generate realistic distribution of emission lines for stellar population models that only provide the continuum.

- 5 + 4 dimensional vector space (continuum + EW)

- K-means clustering to define the continuum classes.

- Test the reconstruction with/without 5+4 dim for clustering and classification.
They propose a simple stochastic recipe to generate realistic distribution of emission lines for stellar population models that only provide the continuum.

- 5 + 4 dimensional vector space (continuum + EW)
- K-means clustering to define the continuum classes.
- Test the reconstruction with/without 5+4 dim for clustering and classification.

Stochastic Recipe for Emission lines
Summary

- They developed an algorithm (noise limited fitting), to accurately measure the emission line parameters (broad, asymmetric emission lines).

- The logEW can be recovered using PCA + LLR in continuum or SDSS broad-band magnitudes.

- Method to generate emission lines for stellar continua of galaxies.

- Further research is necessary for weak lines reconstruction.

- It would account for unknown systematics in spectrophotometry.

- PanSTARRS and LSST

- BPT can be reconstructed only from the continua of galaxies, since one can generate emission lines of galaxy samples only with photometry.
Fim...