Multivariate Classification of Galaxies: which Observables?

D. Fraix-Burnet

Univ. Grenoble-Alpes / CNRS / IPAG

http://ipag.osug.fr/public/fraixbud/

http://astrocladistics.org/
Outline

1. Unsupervised Classification of Galaxies
2. Which Parameters?
3. Phylogenetic Tools
4. Conclusion
Unsupervised Classification of Galaxies

Which Parameters?

Phylogenetic Tools

Conclusion
Unsupervised Classification of Galaxies

Galaxy classification needed for
Galaxy physics and diversification (formation and evolution)
As a probe of their environment – i.e. the Universe (cosmology)

Unsupervised multivariate classification (clustering)
Currently no general, physical, complete and objective classification
Unsupervised Classification of Galaxies

**Multivariate Approaches to Classification in Extragalactic Astronomy**
(Fraix-Burnet, Thuillard, Chattopadhyay 2015, Frontiers in Astronomy and Space Sciences)

- Nearest Neighbor
- Support Vector Machine
- K-Means
- Fuzzy clustering
- Information Bottleneck
- Mixture Models
- Wavelet Analysis
- Hierarchical Classification
- Minimum Spanning Tree (FoF)
- Phylogenetic methods
## Unsupervised Classification of Galaxies

<table>
<thead>
<tr>
<th>Method</th>
<th>$p$</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>kNN</td>
<td>16</td>
<td>colours</td>
</tr>
<tr>
<td>SVM</td>
<td>12</td>
<td>morpho, lumin.</td>
</tr>
<tr>
<td>k-means</td>
<td>1637</td>
<td>spectra</td>
</tr>
<tr>
<td>–</td>
<td>4</td>
<td>$\sigma, \mu_e, r_e, \text{Mg2}$</td>
</tr>
<tr>
<td>–</td>
<td>6 (23)</td>
<td>$\sigma, \mu_e, \text{NaD, D/B, [OIII], [MgbFe]'}$</td>
</tr>
<tr>
<td>–</td>
<td>3</td>
<td>M, r, M/L</td>
</tr>
<tr>
<td>fuzzy partitioning</td>
<td>3</td>
<td>PCA from 8 morph., photom., spectra</td>
</tr>
<tr>
<td>Information Bottleneck</td>
<td>738</td>
<td>spectra</td>
</tr>
<tr>
<td>Mixture Models</td>
<td>4</td>
<td>colours + redshift</td>
</tr>
<tr>
<td>–</td>
<td>10</td>
<td>red. dim. from spectra</td>
</tr>
<tr>
<td>Shapelets</td>
<td>?</td>
<td>images, morpho</td>
</tr>
<tr>
<td>Hierarchical clustering</td>
<td>3</td>
<td>PCA on morphological indicators</td>
</tr>
<tr>
<td>MST</td>
<td>1637</td>
<td>spectra</td>
</tr>
<tr>
<td>Maximum Parsimony</td>
<td>4/6(23)</td>
<td>various</td>
</tr>
</tbody>
</table>
1. Unsupervised Classification of Galaxies
2. Which Parameters?
3. Phylogenetic Tools
4. Conclusion
Which parameters?

Criteria

- availability
  technology evolves with time
- reliability
  uncertainties
- completeness
  missing values, sample diversity
- objectivity
  the parameters affect the physical meaning of the classification
- relevance
  a given parameter set may not be adequate for all clustering tools
Unsupervised Classification of Galaxies

Which Parameters?

Phylogenetic Tools

Conclusion
Diversification of Dark Matter Haloes

Stewart et al. 2008

Fraix-Burnet 2009

Unsupervised Classification of Galaxies
<table>
<thead>
<tr>
<th><strong>Minimum Spanning Tree</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>(also Friends-Of-Friends)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Labelled tree</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>No internal node</td>
</tr>
<tr>
<td>Polynomial time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Semi-labelled tree</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>More general</td>
</tr>
<tr>
<td>NP-hard</td>
</tr>
</tbody>
</table>

\[
w(T) = \sum_e w(e)\]

\[
S = \sum_k \sum_{i>j} d(f_k(i), f_k(j))\]

---

For a L1 norm

\[
w(e) = \sum_k |f_k(i) - f_k(j)|\]

\[
d(f_k(i), f_k(j)) = \sum_k |f_k(i) - f_k(j)|\]

\[
w(T) = \sum_{i>j} \sum_k |f_k(i) - f_k(j)|\]

\[
S = \sum_k \sum_{i>j} |f_k(i) - f_k(j)|\]

\[
\frac{w_{min}}{2} < S_{min} < w_{min}\]
Redundancies

123 Virgo galaxies
48 parameters:
photometry wide and narrow band
line ratios
kinematics
morphology

Linear tree
Dominated by stellar evolution
Incompatibilities, uninformative parameters creating noise, ...

424 galaxies

23 parameters:

- 10 Lick indices: H$\beta$, Fe5015, Mg1, Mg2, Mgb, Fe5270, Fe5335, Fe5406, Fe5709, NaD.
- $[\text{MgbFe}] = \sqrt{\text{Mgb} \ast (0.72 \ast \text{Fe}5270 + 0.28 \ast \text{Fe}5335)}$ indicator of metallicity.
- $\text{Mgb/Fe} = \frac{\text{Mgb}}{(\text{Fe}5270 + \text{Fe}5335)}$ indicator of light-element abundance.
- 6 broad band observables: Brie, Mabs, Kabs, B-R, J-H, and H-K.
- 2 parameters from medium-resolution spectra: $\log \sigma$ and OIII.
- 3 geometrical parameters: $\log(\text{diam})$, logre and D/B.

Fraix-Burnet, Chattopadhyay, Chattopadhyay, Davoust, Thuillard AA, 2012 (hal-00708735)
Objective selection of the parameters
Which parameter subset that can best discriminate different groups?

Principal Component Analysis
axes of greatest variance and correlations between variables

Minimum Contradiction Analysis
best order to arrange the sample on a network or a tree

Maximum Parsimony (Cladistics)
most parsimonious tree representing relationships between groups and objects

Winners: \( \log \sigma \) D/B NaD [\( MgbFe \)]’ Brie Olll
Fraix-Burnet, Chattopadhyay, Chattopadhyay, Davoust, Thuillard AA, 2012 (hal-00708735)
Undocumented parameters ... or too many

WINGS data set \textit{(DFB, M. d’Onofrio, P. Marziani)}

SDSS spectra \textit{(DFB, Bouveyron)}

Unsupervised Classification of Galaxies

D. Fraix-Burnet
1 Unsupervised Classification of Galaxies
2 Which Parameters?
3 Phylogenetic Tools
4 Conclusion
Conclusion

Unsupervised classification of galaxies (and others)

- Exploratory tool
- Insights into data sets and physics
- Needs more investments