

Fast, Linear Time Hierarchical Clustering using the Baire Metric

Pedro Contreras, Fionn Murtagh

(Submitted on 11 Jun 2011)

The Baire metric induces an ultrametric on a dataset and is of linear computational complexity, contrasted with the standard quadratic time agglomerative hierarchical clustering algorithm. In this work we evaluate empirically this new approach to hierarchical clustering. We compare hierarchical clustering based on the Baire metric with (i) agglomerative hierarchical clustering, in terms of algorithm properties; (ii) generalized ultrametrics, in terms of definition; and (iii) fast clustering through k-means partitioning, in terms of quality of results. For the latter, we carry out an in depth astronomical study. We apply the Baire distance to spectrometric and photometric redshifts from the Sloan Digital Sky Survey using, in this work, about half a million astronomical objects. We want to know how well the (more costly to determine) spectrometric redshifts can predict the (more easily obtained) photometric redshifts, i.e. we seek to regress the spectrometric on the photometric redshifts, and we use clusterwise regression for this.

Comments: 27 pages, 6 tables, 10 figures

Subjects: **Machine Learning (stat.ML)**; Information Retrieval (cs.IR); Applications (stat.AP)

MSC classes: 11Z05

ACM classes: H.3.3

Cite as: [arXiv:1106.2229](https://arxiv.org/abs/1106.2229) [stat.ML]

(or [arXiv:1106.2229v1](https://arxiv.org/abs/1106.2229v1) [stat.ML] for this version)

Submission history

From: Fionn Murtagh [[view email](mailto:fionn@cs.cornell.edu)]

[v1] Sat, 11 Jun 2011 12:05:43 GMT (1454kb,D)

[Which authors of this paper are endorsers?](#)

Download:

- [PDF](#)
- [Other formats](#)

Current browse context:

stat.ML

[< prev](#) | [next >](#)

[new](#) | [recent](#) | [1106](#)

Change to browse by:

cs

[cs.IR](#)

stat

[stat.AP](#)

References & Citations

- [NASA ADS](#)

Bookmark([what is this?](#))

