

Pitman-Yor Diffusion Trees

David A. Knowles, Zoubin Ghahramani

(Submitted on 13 Jun 2011 (v1), last revised 16 Jun 2011 (this version, v2))

We introduce the Pitman Yor Diffusion Tree (PYDT) for hierarchical clustering, a generalization of the Dirichlet Diffusion Tree (Neal, 2001) which removes the restriction to binary branching structure. The generative process is described and shown to result in an exchangeable distribution over data points. We prove some theoretical properties of the model and then present two inference methods: a collapsed MCMC sampler which allows us to model uncertainty over tree structures, and a computationally efficient greedy Bayesian EM search algorithm. Both algorithms use message passing on the tree structure. The utility of the model and algorithms is demonstrated on synthetic and real world data, both continuous and binary.

Comments: 8 pages, to be presented at UAI 2011

Subjects: **Machine Learning (stat.ML)**

MSC classes: 62G07, 62H30

ACM classes: G.3.7

Cite as: [arXiv:1106.2494](https://arxiv.org/abs/1106.2494) [stat.ML](or [arXiv:1106.2494v2](https://arxiv.org/abs/1106.2494v2) [stat.ML] for this version)

Submission history

From: David Knowles [[view email](#)]

[v1] Mon, 13 Jun 2011 17:23:50 GMT (1059kb,D)

[v2] Thu, 16 Jun 2011 11:06:04 GMT (1059kb,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:

[stat](#)

References & Citations

- [NASA ADS](#)

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

