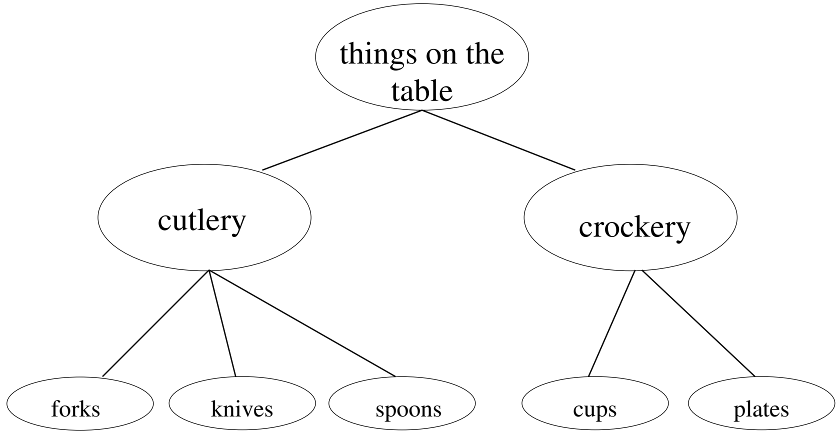
Hierarchy and Object Categories

Chris Williams

School of Informatics, University of Edinburgh, UK

Motivation



The world has hierarchical structure, e.g.:

- Grouping of organisms into families, genera, species, sub-species etc.
- Human artifacts

- A challenge: given observations of many human artifacts, could we create a meaningful hierarchical clustering?

- IS-A and PART-OF hierarchical structure

- Is a tree structure rich enough/appropriate?

Overview

- Extracting an Ontology of Portrayable Objects from Wordnet
- Hierarchical Agglomerative Clustering (HAC)
- Hierarchical Mixture Modelling (Williams, 2000)
- Dirichlet Diffusion Trees (Neal, 2001)
- Hierarchical Topic Models (Blei et al, 2004)
- Autoclass IV (Hanson, Stutz and Cheeseman, 1991)
- Finding objects in images: PASCAL dataset

Extracting an Ontology of Portrayable Objects from Wordnet

- S. Zinger, C. Millet, B. Mathieu, G. Grefenstette, P. Hède, P.-A. Moëllic
  MUSCLE/ImageClef Workshop on Image and Video Retrieval Evaluation, Sept 2005.
- WordNet, `wordnet.princeton.edu`
- Build an ontology using the *hyponym/hypernym* relationship
- Their upper level ontology has 102 nodes (some hand pruning involved)
- There are over 24,000 terms corresponding to objects in Wordnet
- object $\rightarrow$ {living thing, natural object, artifact, floater, ice, web }

- object

↓ artifact

↓ instrumentality

↓ conveyance

↓ vehicle

↓ wheeled vehicle

↓ self-propelled vehicle

↓ motor vehicle

- car

motor vehicle, direct hyponyms:

- amphibian, amphibious vehicle
- bloodmobile
- car, auto, automobile, machine, motorcar
- doodlebug (a small motor vehicle)
- four-wheel drive, 4WD
- go-kart
- golfcart
- hearse
- motorcycle

- snowplough
- truck

Hierarchical clustering

Calculate all pairwise distances d_{ij}, e.g.

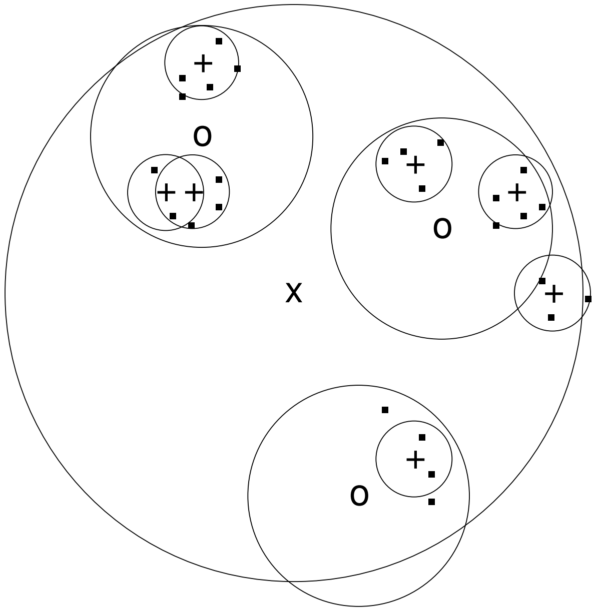
$$d_{ij}^2 = |\mathbf{x}_i - \mathbf{x}_j|^2$$

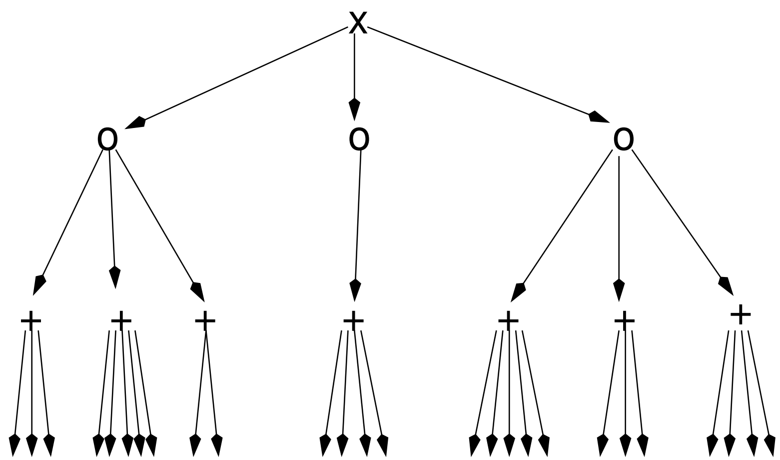
- Find the closest pair of points
- Replace this pair with a single new "clump"
- Repeat, until some suitable stopping criterion is met

- Results can be displayed as a *dendrogram*
- This is *agglomerative* clustering; divisive techniques are also possible

```
                             _____________|------> p08
                            |             |______|------> p04
                            |                    |------> p09
 |--------------------------|                       _______|----> p02
 |                          |             |--------|       |----> p12
 |                          |-------------|        |-------> p14
 |                                        |________|------> p10
-|                                                 |------> p15
 |                                          ________|-----> p03
 |                          |--------------|        |-----> p06
 |                          |              |         ________|-----> p01
 |--------------------------|              |--------|        |-----> p07
                            |                       |--------> p11
                            |______________|-------> p05
                                           |_______|------> p13
                                                   |______|-----> p16
                                                          |-----> p17
```

Williams (2000): The basic idea





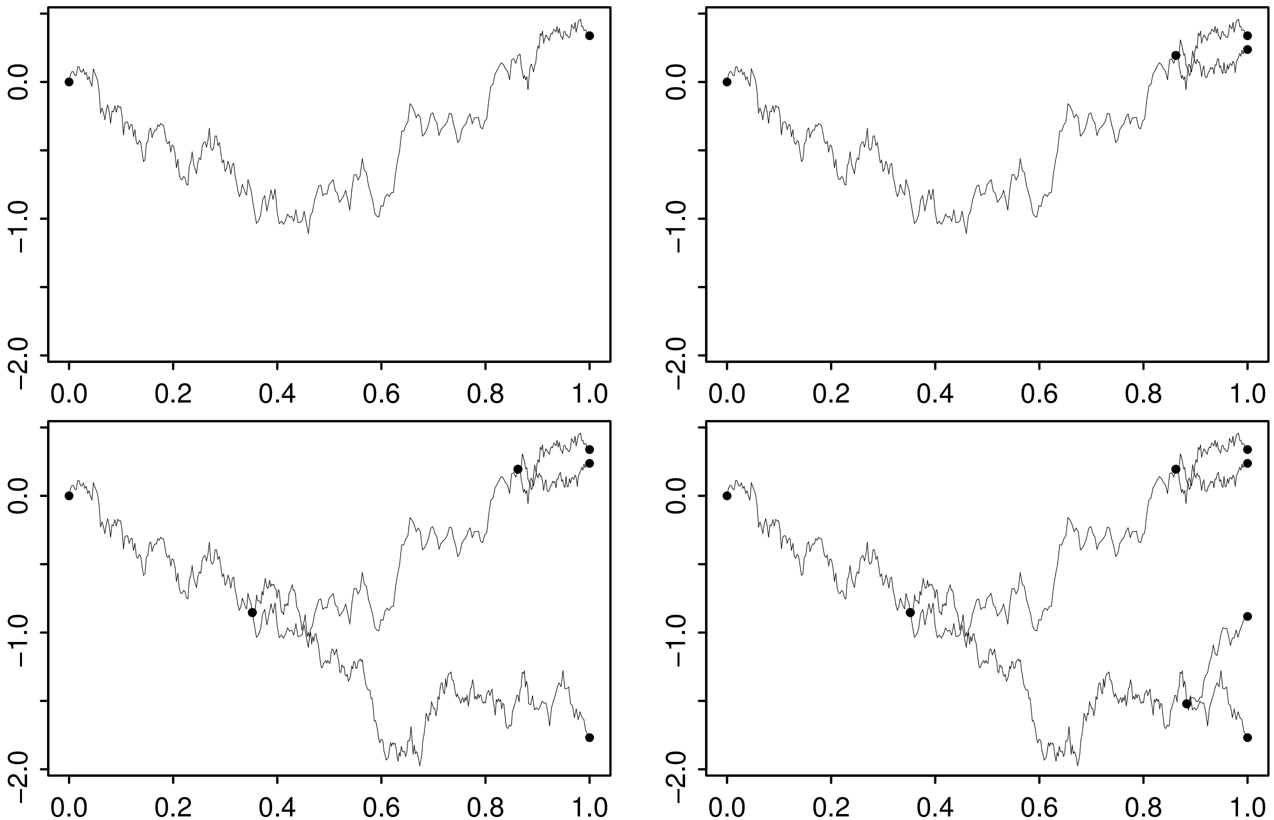
- A leaf variable is a sum of RVs along the path from root to leaf
- Number of units: Markovian model $P(\mathbf{n}) = P(n_1)P(n_2|n_1)\dots P(n_L|n_{L-1})$
- Prior over trees: each unit can connect to any parent in the layer above with equal probability.
- $P(X|\mathtt{tree\ structure})$ can be calculated by noting that this is a branching Gaussian process
- Inference using MCMC
- Limitation: finite mixture model

Dirichlet Diffusion Trees

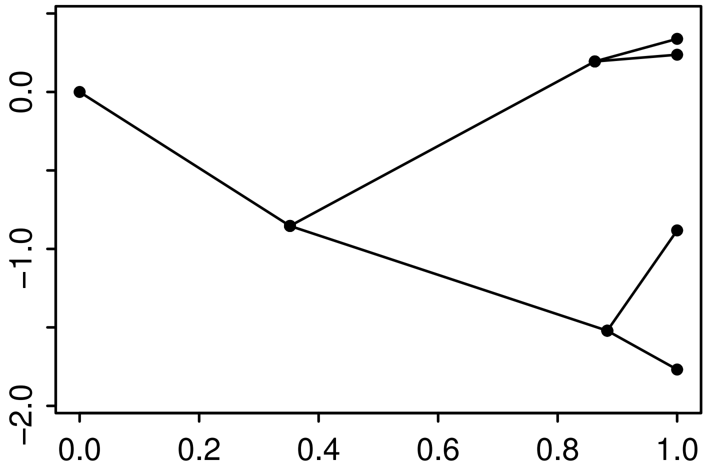
(Based on Radford Neal's slides, NTOC 2001)

- Dirichlet process mixtures have a countably infinite number of components, but do not capture hierarchical structure. Dirichlet Diffusion Trees are designed to overcome this
- To generate the first point: simulate Brownian motion from $t = 0$ to $t = 1$
- To generate the next point: Simulate Brownian motion again, but follow the first path initially, diverging at some random time
- To generate later points: Follow previous paths initially, choosing randomly when they branch, but diverge at some random time. Paths followed many times previously are more likely to be followed again
- This produces an *exchangeable* distribution over data points
- The probability of divergence from previous paths is controlled by the *divergence function*

Samples 1 to 4



Suppressing the details



Sample from the posterior using MCMC

Has been applied to clustering gene expression data

Phylogenetic Trees

- Discrete variables (genes) with mutation CPTs that depend on the elapsed time

- Infer tree structure and times (MCMC)

- Reference: Joseph Felsenstein, *Inferring Phylogenies*, Sinauer (2004)

Hierarchical Topic Models

(Blei, Griffiths, Jordan Tenenbaum; 2004)

Generate a document as follows:

- At each node in the tree have a distribution over words $p(w|z, \beta)$
- Choose a path from the root to a leaf of a L-level tree
- Draw a vector of topic proportions θ from an L-dimensional Dirichlet distribution
- Generate words in the document from the mixture of topics from root to leaf

$$p(w|\theta) = \sum_{i=1}^{L} \theta_i p(w|z = i, \beta_i)$$

- This describes a hierarchical topic model wrt a fixed tree. A Chinese Restaurant Process is used to place a prior over possible trees
- Inference is carried out using Gibbs sampling

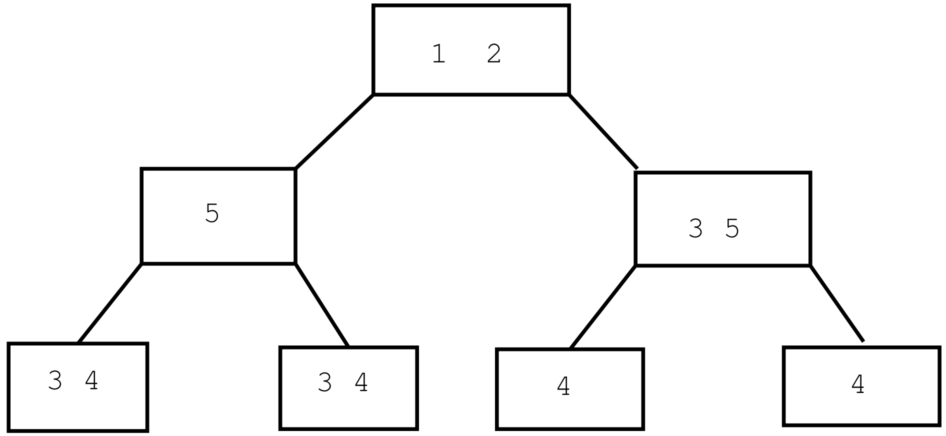
Bayesian Classification with Correlation and Inheritance

Hanson, Stutz and Cheeseman (IJCAI, 1991)

- Variables in each mixture component are divided into *blocks*: full modelling within blocks, independence between blocks

- The block of attributes that are irrelevant is shared by all components

- In general arrange components in a tree structure: each block is placed at some level in the tree, to be shared by all nodes below



- Autoclass IV developed around the full hierarchical block structure

- See also Vaithyanathan and Dom (ICML, 1999) for application to document clustering

Summary

In probabilistic hierarchical clustering the leaf entities share properties with other leaves depending on the common path

- For Dirichlet diffusion trees one *adds* RVs along the path

- For the hierarchical topic model the combination mechanism is a *mixture* distribution of topics along the path

- For Autoclass IV nearby leaves share more blocks of parameters

PASCAL VOC Challenge second edition

Andrew Zisserman, Chris Williams

- First challenge, see `http://www.pascal-network.org/challenges/VOC/`
- Classification and localization tasks
- Likely classes: cars, motorbikes, bicycles, people, cows, sheep, cats, dogs, trucks, buses
- mid Jan: release development set (training/validation data)
- mid-end Feb: release test set
- mid March: results deadline

References

- S. Zinger, C. Millet, B. Mathieu, G. Grefenstette, P. Hède, P.-A. Moëllic. Extracting an Ontology of Portrayable Objects from Wordnet. MUSCLE/ImageClef Workshop on Image and Video Retrieval Evaluation, Sept 2005.

- C. K. I. Williams.. A MCMC approach to Hierarchical Mixture Modelling. In Advances in Neural Information Processing Systems 12, eds. S. A. Solla, T. K. Leen and K-R. Muller, MIT Press (2000) .

- Neal, R. M. Defining priors for distributions using Dirichlet diffusion trees, Technical Report No. 0104, Dept. of Statistics, University of Toronto (2001). [see also R. M. Neal. Density modeling and clustering using Dirichlet diffusion trees, in J. M. Bernardo, et al. (editors) Bayesian Statistics 7, pp. 619-629, (2003).]

- D. M. Blei, T. L. Griffiths, M. I. Jordan, J. B. Tenenbaum. Hierarchical Topic Models and the Nested Chinese Restaurant Process. In Advances in Neural Information Processing Systems 16, eds. S. Thrun, L. Saul and B. Schölkopf, MIT Press (2004).

- R. Hanson, J. Stutz, P. Cheeseman. Bayesian Classification with Correlation and Inheritance. In Proc. Twelfth International Conference of Artificial Intelligence (IJCAI 91), Morgan Kaufmann (1991).