Probabilistic Modeling using Tree Linear Cascades

Nick Landolfi & Sanjay Lall
Stanford University

Revision: 14 April 2022

Motivation

- want to model *functional or causal relationships among high-dimensional multivariate data*
  - control systems, genomics, cloud telemetry, fMRI brain imaging, ...
- for the purposes of tasks like *classification* or for *interpretation* of phenomena
  - want *parsimonious representations* of, i.e. $O(d)$, of high-dimensional continuous distributions
  - want *computationally attractive* techniques for finding these representations
- this paper
  - formulates and solves the *simultaneous cascade regression* fitting problem
    - finds a linear and tree-structured structural equation model (SEM)
  - analyzes *tree linear cascades* connecting them to the regression
    - these are a particular SEM with *neither Gaussian nor independent assumptions* on errors
  - connects both to the classical Chow-Liu result for Gaussian densities

Simultaneous cascade regression

- want to *simultaneously search for graph and functional relations* of structural equation model
- **Problem 1**. Find rooted tree (T, r) on $\{1, 2, \ldots, d\}$ and $A \in \mathbf{R}^{d \times d}$ to

$$\begin{array}{ll} \text{minimize} & \mathbf{E}\,||Ax - x|| \\ \text{subject to} & A \in \mathtt{sparse}(T, r). \end{array}$$

  - where $\mathtt{sparse}(T, r) = \{A \in \mathbf{R}^{d \times d} \mid A_{ij} = 0 \text{ if } j \neq \mathrm{pa}_i\}$,
  - has elements with same *sparsity pattern as the directed adjacency matrix* of (T, r)
  - makes *no distributional assumptions* on x
- **Solution.** Find maximum spanning tree of complete graph with edges weighted by *squared correlations*
  - and for selected edge $\{i, j\}$ with $j = \mathrm{pa}_i$, choose $A_{ij} = \mathbf{E}(x_i x_j)$
  - see Lemma 1 and Theorem 2 of paper

Tree linear cascades

- let (T, r) be a rooted tree on $\{1, 2, \dots, d\}$; given
    1. an uncorrelated random vector $e : \Omega \to \mathbf{R}^d$ with $\mathbf{E}(e) = 0$ and $\mathrm{cov}(e) \succ 0$ and
    2. a matrix $A \in \mathrm{sparse}(T, r) = \{A \in \mathbf{R}^{d \times d} \mid A_{ij} = 0 \text{ if } j \neq \mathrm{pa}_i\}$
- x is a *tree linear cascade* on e with respect to A if

$$x = Ax + e$$

- **Result**: T is the *unique* maximum spanning tree of graph with edges weighted by *squared correlations*
    - i.e., *T is identifiable* from distribution of x; the root is *not*; see Section III.B of paper
    - consequently, *cascade regression identifies the tree* of a tree linear cascade
    - see Theorem 1 and Corollary 1 of paper

Generalization of Gaussian Chow-Liu

- **Problem 2.** Given a density $g : \mathbf{R}^d \to \mathbf{R}$, find a tree T on $\{1, \dots, d\}$ and a density $f : \mathbf{R}^d \to \mathbf{R}$ to

$$\begin{array}{ll} \text{minimize} & d_{kl}(g, f) \\ \text{subject to} & f \text{ factors according to } T \end{array}$$

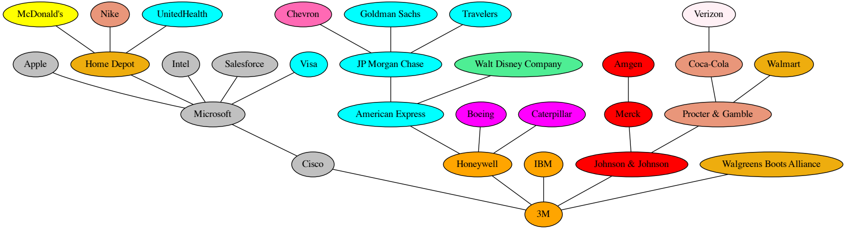
  - density case of classical Chow-Liu problem; d_{kl} is the Kullback-Leibler divergence
  - *well-known* solution: maximum spanning tree of complete graph with edges weighted by *mutual informations*
- **Connection.** Gaussian Chow-Liu and cascade regression *trees coincide*; see Corollary 2
  - however, *cascade regression makes no Gaussian assumption*
    - not all tree linear cascades are tree Gaussians...
    - but converse is true, use sparse Cholesky factorization; see Section III.A of paper
  - consequently, cascade regression provides *alternate justification/interpretation of Chow-Liu tree*
    - conversely, suggests interpreting *cascade regression as density approximation*

Empirical cascade regression: a simple stock example

- in practice, we have data $x^1, \ldots, x^n \in \mathbf{R}^d$ and we want to find (T, r) and A

$$\text{minimize} \quad \sum_{k=1}^{n} \|Ax^k - x^k\| \quad \text{subject to} \quad A \in \text{sparse}(T, r)$$

- ten years of daily stock price data from the wall street journal for the Dow Jones 30, here's the tree



  - nodes are stocks colored by industry; roughly speaking *stocks in similar industries are connected*
  - dataset and straightforward code available online

Conclusion

- the paper provides *several theoretical results*
  - posing and solving cascade regression
  - analyzing tree linear cascades
  - giving a non-Gaussian interpretation of Chow-Liu
- these results lead to *computationally attractive* practical techniques
- next steps may include other applications, or other problem variants
  - e.g., block case; extending prior work on stochastic process case (see [Materassi & Innocenti 2010])
- more details and full proofs available in paper and at poster session, *thanks!*