Discovering Hidden Variables in Noisy-or Networks

Yacine Jernite, Yoni Halpern, David Sontag
New York University

Noisy-or Networks

- Each latent variable is drawn $Y_i \sim \text{Bernoulli}(\theta_i)$.
- Each latent variable with $Y_i = 1$ activates each observed variable $X_j$ with probability $1 - f_{ij}$.
- An observed variable is "on" ($X_j = 1$) if any parent activates it.

Singly-coupled Observations

- A set of observed variables $S = \{X_1, \ldots, X_n\}$ is singly-coupled if the following two conditions hold:
  1. There exists a coupling parent, $Y_i$, which is a parent of all $X_j \in S$.
  2. There is no other parent which is shared by any pair of $X_j, X_k \in S$.
- Marginally, $S$ is a binary mixture model. When $|S| = 3$, there is an analytical decomposition to estimate parameters $p_i$ and $f_{ij}$ for all $X_j \in S$ [2, 3].

Discovering Hidden Variables

- Without structure, how do we determine whether a triplet is singly-coupled? It’s impossible! Whether a triplet is singly-coupled is not identifiable.
- A rank test (flattened rank) can determine whether quartets are singly-coupled.
- When a singly-coupled quartet is found, we can take the following steps:
  1. Learn parameters, with singly-coupled triplets.
  2. Extend the hidden variable to its other children and learn those parameters as well.

Extending Hidden Variables

- Define pointwise mutual information (PMI) and conditional pointwise information (CPI) between variables as:
  $$\text{PMI}(X_1, X_2) = \frac{P(X_1, X_2)}{P(X_1)P(X_2)}$$
  $$\text{CPI}(X_1, X_2 | X_3) = \frac{P(X_1, X_2 | X_3)}{P(X_2 | X_3)}$$
- Extending test: Let $(X_1, X_2)$ be observable variables singly-coupled by $Y_i$, and $X_3$ be a variable whose relationship with $Y_i$ is unknown.
- \(\text{CPI}(X_1, X_2 | X_3) \leq \text{PMI}(X_1, X_2)\) with equality if and only if $X_3$ is not a child of $Y_i$.
- The ratio of CPI $(X_1, X_2 | X_3) \text{ and PMI}(X_1, X_2)$ can be used to solve for $f_{ij}$. Solution uses estimates of $p_i, f_{ij}, f_{ij}$ obtained from step 1 above.

Identifiability & Guarantees

- A network is called $\epsilon$-quartet learnable if:
  1. Every latent variable has a singly-coupled quartet (possibly after subtraction).
  2. All quartets that are not singly-coupled have an unfolding with a third eigenvalue of at least $\epsilon$. (Holds with high probability for randomly chosen parameters.)
- Our algorithm learns $\epsilon$-quartet learnable networks of constant depth with polynomial sample complexity. Dependence on depth is exponential (often small).

Structure learning algorithm

- The STRUCTURE-LEARN algorithm uses the rank test to discover hidden variables and the extending step to find all of their children. A subtraction step [1] allows us to create more singly-coupled variables by subtracting the influence of learned parents.
- Networks can be characterized by the required number of serial subtractions (depth).
- We learn nearly all of the $\text{QMRF-DT}$ (Quick Medical Reference) network [4] for medical diagnosis!

Experiments

- We present a qualitative comparison to the variational EM method from [5] using the same toy image network from the paper.

Conclusions

- We have shown that it is possible to learn the structure and parameters of noisy-or bipartite networks with polynomial sample complexity and provable guarantees.
- Future work includes:
  - Demonstrating robustness in real world applications.
  - Expanding the class of non-linear link functions that can be learned in this way.

References


Figures:
- Figure 1: When structure is known, we can identify singly-coupled triplets and use them to learn the model parameters using the method of moments, as in [3].
- Figure 2: Learning parameters from a noisy-or network.
- Figure 3: Learning parameters from a noisy-or network.