Discovering Hidden Variables in Noisy-or Networks

Yacine Jernite, Yoni Halpern, David Sontag

New York University

Noisy-or Networks

Generative model:

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

Singly-coupled Observations

- A set of observed variables $\mathcal{S} = \{X_1, \ldots, X_{|\mathcal{S}|}\}$ is singly-coupled [1] if the following two conditions hold:

1. There exists a **coupling parent**, Y_i, which is a parent of all $X_j \in \mathcal{S}$.
2. There is **no other parent** which is shared by any pair of $X_j, X_k \in \mathcal{S}$.

- Marginally, $\mathcal{S}$ is a binary mixture model. When $|\mathcal{S}| = 3$ there is an analytical decomposition to estimate parameters p_i and $f_{i,j}$ for all $X_j \in \mathcal{S}$ [2, 3].

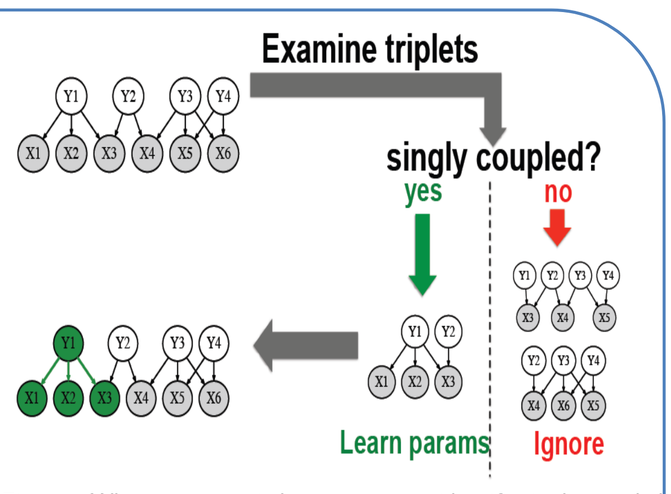
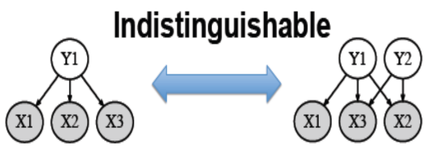


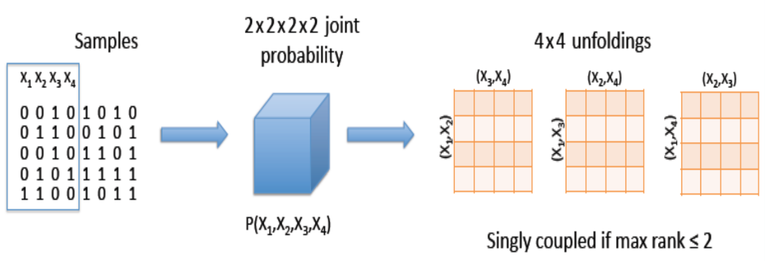
Figure : When structure is known, we can identify singly-coupled triplets and use them to learn the model parameters using the method-of-moments, as in [1].

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, as with singly-coupled triplets.
2. Extend the hidden variable to find its other children and learn those parameters as well.

Extending Hidden Variables

- Define pointwise mutual information (PMI) and conditional pointwise information (CPMI) between variables as:

$$\text{PMI}(X_1, X_2) = \frac{P(\bar{X}_1, \bar{X}_2)}{P(\bar{X}_1)P(\bar{X}_2)}$$

$$\text{CPMI}(X_1, X_2 | X_3) = \frac{P(\bar{X}_1, \bar{X}_2 | \bar{X}_3)}{P(\bar{X}_1 | \bar{X}_3)P(\bar{X}_2 | \bar{X}_3)}$$

Extending test: *Let (X_1, X_2) be observable variables singly-coupled by Y_i and X_j be a variable whose relationship with Y_i is unknown.*

$$\text{CPMI}(X_1, X_2 | X_j) \leq \text{PMI}(X_1, X_2),$$

with equality if and only if X_j is not a child of Y_i.

- The ratio of $\text{CPMI}(X_1, X_2 | X_j)$ and $\text{PMI}(X_1, X_2)$ can be used to solve for $f_{i,j}$. Solution uses estimates of $p_i, f_{i,1}, f_{i,2}$ obtained from step 1 above.

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.

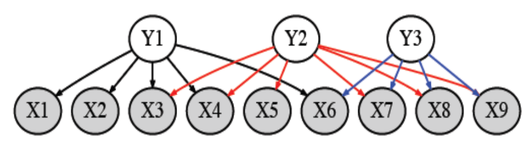


Figure : $\{X_1, X_2, X_3, X_6\}$ is singly-coupled by Y_1. The extend step finds X_4, the final child of Y_1. $\{X_3, X_4, X_5, X_7\}$ is singly-coupled by Y_2 *after subtracting off* Y_1, and so on. The structure of this network and its parameters are entirely learnable by our algorithm.

- Networks can be characterized by the required number of serial subtractions (depth).
- We learn nearly all of the aQMR-DT (Quick Medical Reference) network [4] for medical diagnosis!

depth	diseases discovered	edges learned
0	469	39,522
1	82	4,875
2	13	789
3	2	86
inf	4	198

- What's left:

Identifiability & Guarantees

- A network is called ϵ-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 ϵ. (Holds with high probability for randomly chosen parameters.)

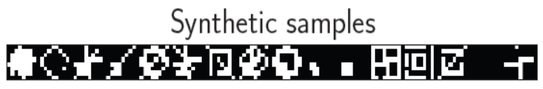
- Our algorithm learns ϵ-quartet learnable networks of constant depth with polynomial sample complexity. Dependence on depth is exponential (often small).

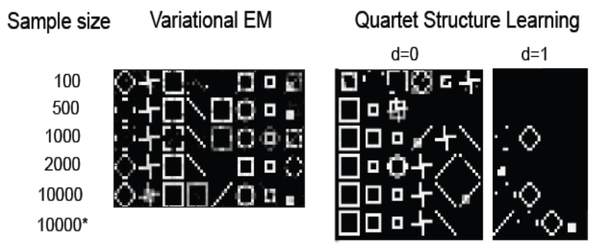
Experiments

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

Hidden sources:

Synthetic samples





The method of moments approach runs approximately 50x faster on the 10K samples.

Conclusion

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

[1] Y. Halpern and D. Sontag. "Unsupervised Learning of Noisy-Or Bayesian Networks". In: *Conference on Uncertainty in Artificial Intelligence (UAI-13)*. 2013. [2] P. Lazarsfeld. "Latent Structure Analysis". In: *Measurement and Prediction*. Ed. by S. Stouffer et al. Princeton, New Jersey: Princeton University Press, 1950. [3] J. Pearl and M. Tarsi. "Structuring causal trees". In: *Journal of Complexity* 2.1 (1986). [4] M. A. Shwe et al. "Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base". In: *Meth. Inform. Med* 30 (1991). [5] T. Šingliar and M. Hauskrecht. "Noisy-or component analysis and its application to link analysis". In: *The Journal of Machine Learning Research* 7 (2006).