Neural Networks 103 (2018) 44–54

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

# A multivariate additive noise model for complete causal discovery

Pramod Kumar Parida [a,*], Tshilidzi Marwala [a], Snehashish Chakraverty [b]

[a] Electrical & Electronic Engineering Science, University of Johannesburg, South Africa
[b] Department of Mathematics, National Institute of Technology Rourkela, Odisha, India

ARTICLE INFO

*Article history:*
Received 13 May 2017
Received in revised form 8 March 2018
Accepted 16 March 2018
Available online 26 March 2018

*Keywords:*
Additive noise models
Causal independence
Causal influence factor
Model fitting error

ABSTRACT

Explaining causal reasoning in the form of directed acyclic graphs (DAGs) yields nodal structures with multivariate relationships. In real-world phenomena, these effects can be seen as multiple feature dependency with unmeasured external influences or noises. The bivariate models for causal discovery simply miss to find the multiple feature dependency criteria in the causal models. Here, we propose a multivariate additive noise model (MANM) to solve these issues while analyzing and presenting a multi-nodal causal structure. We introduce new criteria of causal independence for qualitative analysis of causal models and causal influence factor (CIF) for the successful discovery of causal directions in the multivariate system. The scores of CIF provide the information for the goodness of casual inference. The identifiability of the proposed model to discover linear, non-linear causal relations is verified in simulated, real-world datasets and the ability to construct the complete causal model. In comparison test, MANM has out performed Independent Component Analysis based Linear Non-Gaussian Acyclic Model (ICA-LiNGAM), Greedy DAG Search (GDS) and Regression with Sub-sequent Independent Test (RESIT), and performed better for Gaussian and non-Gaussian mixture models with both correlated and uncorrelated feature relations. In performance test, different model fitting errors which occur during causal model construction are discussed and the performance of MANM in comparison to ICA-LiNGAM, GDS and RESIT is provided. Results show that MANM has better causal model construction ability, producing few extra sets of direction with no missing or wrong directions and can estimate every possible causal direction in complex feature sets.



## 1. Introduction

The dependency in different factors raises the causal reasoning in real-world processes. The simplest structures defined by Directed Acyclic Graphs (DAGs) are primarily analyzed using Structural Equation Models (SEMs) (Bollen, 1989). Later a more complex system Bayesian Network (BN) (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993) was designed to provide structural construction conditions using probability criteria for dependent features. Bayesian analysis only provides the functionality to construct DAGs using Bayesian low, graph d-separation, Markov Equivalent Classes, v-structures, Markov Blanket and Markov Chain Monte Carlo (MCMC) simulation. These assumptions lead to a model where nodal relations are defined only using the directed edges and weights of connection strengths to show the amount of information passed on. By definition, the works carried out on causal analysis only establish the model in the form of directed acyclic graph with weighted values on the edge to find the causal model. But none of these inform the quality of causal inference.

Most causal models such as Geiger and Heckerman (1994), Pearl (2000) and Spirtes et al. (1993) use Gaussian assumption of the data to provide causal reasoning using Markov Equivalent Classes. Non-Gaussian properties of data can solve the problem for better approximation over real-world cases, and so the model of Linear Non-Gaussian Acyclic Model (LiNGAM) was proposed by Shimizu, Hyvärinen, Kano, and Hoyer (2005) for causal discovery. The LiNGAM works with Independent Component Analysis (ICA) to estimate causal ordering by measuring the effects of different components independently. If it is known how one feature affects the other, then the causal direction can be found. The successive arrangement of causal directions, starting from the most independent node to the most dependent node, provides causal ordering. But in LiNGAM, they first estimate the order and then from the order, they find the causal directions, which is an incorrect process. In addition the later developments over LiNGAM like Shimizu, Hoyer, Hyvärinen, and Kerminen (2006) and Shimizu et al. (2011) also get it wrong.

The nonlinear causal discovery using the additive noise model by Hoyer, Janzing, Mooij, Peters, and Schölkopf (2009) estimates

* Corresponding author.
*E-mail addresses:* pramodsstyle@gmail.com (P.K. Parida), tmarwala@gmail.com (T. Marwala), sne_chak@yahoo.com (S. Chakraverty).

https://doi.org/10.1016/j.neunet.2018.03.013
0893-6080/

causal directions in two variables using nonlinear functions over joint probability densities of the independent variable, added with arbitrary noise. The bivariate model is capable of finding directions by comparing two variables at a time. The problem with the bivariate model is in the accuracy of estimation, when a node or feature is dependent on multiple other features, that is if multiple nodes have causal directions towards a single node then the model estimates wrong directions. Causal discovery for the continuous case was given in Peters, Mooij, Janzing, and Schölkopf (2014) using the additive noise model. This model relies on the independence of residuals while observed nodes are regressed on each other. The dependency is measured using the Hilbert–Schmidt independence criterion by taking minus log over the functions of the feature set. Although the method claims to detect all possible edges in the model, the results show that method fails mostly for multivariate data with large node sets. The additive noise model for cyclic causal models was proposed by Mooij, Janzing, Heskes, and Schölkopf (2011). But they have only considered the case for bivariate model and their system is restricted to bivariate Gaussian-noisy models.

All the proposed works on bivariate analysis fail to hold on to the graph d-separation condition. The d-separation uses triplet node sets (a set of 3 nodes at least) to observe the flow of information between parent and child nodes by blocking connected paths. While using a bivariate model, the d-separation criteria cannot be verified, which is a fundamental and essential condition for defining the causal models. So, an effective and easier method of multivariate analysis is proposed in this paper, to find the complete causal structure using a multivariate additive noise model. It shows that the proposed method is fully capable of constructing multi-nodal causal structures for complete causal analysis. The insufficiency of the bivariate model which leads to the development of MANM is discussed, to provide the necessity of the proposition.

The conditional independence for causal inference is not one of the best tools for the estimation of independent features in the observational set. While the findings of causal inference are about qualitative analysis, the use of conditional independence/dependence merely provides information on quantitative measures. As a solution to this, the causal independence is introduced which is better suited for causal inference and analysis. The major difference between conditional independence and causal independence is discussed with the mathematical explanation to validate the claim.

Until now, the proposed methods emphasize the computation of directed edges from connection strength values. The connection strength is a quantification of the information that flows from a parent node to child node. On the other hand, the causal inference is all about retrieving the influence of the information which causes it. Also, the issue of finding the goodness of causal inference has not been addressed or pointed out yet. Only finding the values for connection strengths which explicitly contribute towards the measurement for the amount of information passed on, does not conclude anything regarding the quality of inference. In this paper for the first time, the new concept of the Causal Influence is introduced, which provides the measure for the goodness of causal inference. The value of causal influence is crucial to detect the causal direction based on its influence on the child nodes. The maximized causal influence factor (CIF) values are used for the successful discovery of causal directions in the feature set when the substructures are d-separable V-structures.

While most of the proposed methods do not scope for the analysis of mixture models of Gaussian and non-Gaussian distributions, the proposed MANM can handle these datasets with ease. To show the capability of MANM method, the complexity of the mixture models has been increased by making them both correlated, uncorrelated and mixed cases of both in simulation tests. The proposed model (MANM) is compared with Independent Component Analysis based Linear Non-Gaussian Acyclic Model (ICA-LiNGAM) (Shimizu et al., 2006), Greedy DAG Search (GDS) (Chickering, 2002) and Regression with Sub-sequent Independent Test (RESIT) (Peters et al., 2014) over simulated models and the test results are provided in Section 5.3. The choice of these three methods is due to their major contributions and wide usability in causal inference studies. Also, these all provide maximum inference on any dataset of interest and can handle large node sets.

The main goal of causal analysis is to find the directionally connected acyclic nodal structures which can be used for complete causal explanation. While constructing such models, many problems arise, due to the wrongly estimated directions which result in missing directions, and mostly due to overestimated extra sets of directions. These issues are studied in comparison with ICA-LiNGAM, GDS, and RESIT, to find the effectiveness and accuracy of MANM while constructing causal models from the estimated causal directions. Results for model under-fitting, reverse-fitting and over-fitting are calculated using mathematical formulas and graphs are provided to show the performances of MANM in comparison with other methods.

We start our discussion by defining the multivariate additive noise model in Section 2, and the identifiability of the model towards complete causal estimations is discussed in Section 3. Section 4 provides the tools for estimation using the proposed method and experimental results for simulated causal structures, and real-world tests are provided in Section 5. The performance of the proposed method is verified with ICA-LiNGAM, GDS and RESIT, for their model constructibility and accuracy in Section 5.4. A concise discussion of the work and future perspective are given in Section 6.

## 2. Multivariate additive noise model: MANM

The proposed multivariate additive noise model (MANM) depends on the imposed system assumptions to produce a complete causal model in the form of Directed Acyclic Graphs (DAGs). The assumptions are as follows for MANM: Consider the multi-nodal structure $G$ as a DAG with $n$ number of nodes represented by $\{X_i\}$ where $(i = 1, \ldots, n)$ and each $X_i$ is a matrix of $(m \times 1)$. All the nodes in $G$ are arranged in order from the most causal independent node at the top of the graph to the least causal independent ones towards the bottom. So, the edge directions in the DAG $G$ are from the top towards the bottom, and none of the nodes observed later than the earlier ones have directed edge towards any earlier node. The above assumptions produce a causal model where the graph starts from the parent node and goes down to descendant nodes while following a **top-bottom** graphical construction.

*Causal independence.* Any causal direction $\{x_i \to x_j\}$ can be described using the causal order $O(x_i) > O(x_j)$. Consider a child node $x_h$ and its parent node set $Pa(x_h) = \{x_i, x_j, x_k, x_l\}$, where parents are of the following order $\{O(i) > O(j) > O(k) > O(l)\}$. Here $x_h$ (child node) can only be influenced by the parent set $Pa(x_h)$ ($x_i, x_j, x_k, x_l$ are parents of $x_h$) through the information transferred from the parent nodes. The direction $\{Pa(x_h) \to x_h\}$ makes $Pa(x_h)$ to become causally independent of $x_h$ i.e. $x_h$ is causally dependent on $Pa(x_h)$ as the information flows from parent to child node. Whereas both $Pa(x_h)$ and $x_h$ are conditionally dependent on each other, which is a quantitative information and not useful for causal analysis. The next section provided a more detailed explanation of this. Notice that every node in $Pa(x_h)$ has directed edges towards $x_h$, but the converse is not true. In the parent set node $x_i$ is most causal independent and $x_l$ is least causal independent (as from the order set).

While observing datasets, the variables are represented as nodes in the causal model with multivariate relations. Assume