CAUSALITY FOR REMOTE SENSING: AN EXPLORATORY STUDY

Soronzonbold Otgonbaatar, Mihai Datcu, Begüm Demir

DLR Oberpfaffenhofen, Technical University of Berlin

ABSTRACT

Causality is one of the most important topics in a Machine Learning (ML) research, and it gives insights beyond the dependency of data points. Causality is a very vital concept also for investigating the dynamic surface of our living planet. However, there are not many attempts for integrating a *causal model* in Remote Sensing (RS) methodologies. Hence, in this paper, we propose to use patch-based RS images and to represent each patch-based image by a single variable (e.g. entropy). Then we use a Structural Equation Model (SEM) to study their cause-effect relation. Moreover, the SEM is a simple causal model characterized by a Directed Acyclic Graph (DAG). Its nodes are causal variables, and its edges represent causal relationships among causal variables if and only if causal variables are dependent.

***Index Terms*—** causality, machine learning, earth observation, remote sensing

1. INTRODUCTION

The probability distribution of data points in RS datasets can be represented by a Bayesian graph network [1], [2]. The Bayesian graph network is a graph network factorized according to a Bayesian probability distribution, and its nodes represent data variables and its edges are interactions among these data variables. Moreover, intervened data variables are ubiquitous in RS data points due to the constant changes in Earth's surface, and this intervention leads to an intervened probability distribution. The intervened probability distribution conveys causal relations (cause-effect relations) among the RS data points. However, the Bayesian graph network does not provide much information on the intervened data points. Thus, some studies proposed a causal graph network for identifying and analysing causal relations of dependent data points [3], [4]. A causal graph network is a directed acyclic graph for data points: data points preside at its nodes, and its directed edges characterize their causal relations. In this work, we focus on how to represent RS datasets on the nodes of a causal graph network, and how to integrate a causal graph model in RS methodologies. In addition, we assume that the causal relation of a given RS dataset is known to avoid discovering them [5], [6] which is an intractable problem for classical algorithms given n nodes. Discovering a causal relation refers to finding directed edges among the nodes.

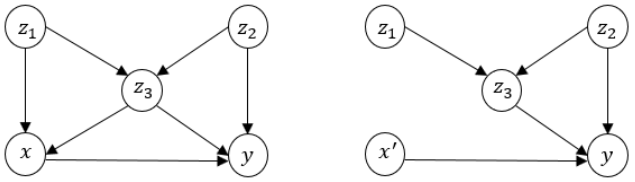

Fig. 1. [Left] An DAG G, [Right] An intervened (truncated) $do(X_x = X_{x'})$ DAG $\bar{G}$.

2. CAUSAL DIRECTED ACYCLIC GRAPH

A causal graph network is a directed acyclic graph (DAG): A directed graph G consists of vertices $V \in \{x, y, z_1, z_2, z_3\}$ and edges E (see Fig. 1 [Left]). If all edges E are directed ($z_1 \to x$, $x \to y$, etc.) then the graph is called a *directed* graph. A *directed path* from z_1 to y is a path between z_1 and y where all edges are pointing towards y: $z_1 \to x \to y$ and $z_1 \to z_3 \to y$. A *cycle* is a path $(z_1, x, \ldots, y)$ with an additional edge between z_1 and y, and a *directed cycle* is a directed path $(z_1, x, \ldots, y)$ from z_1 to y with an edge $y \to z_1$. Moreover, the DAG is a directed graph without the *directed cycle* as illustrated in Fig. 1 [Left]. For simplicity, we name a causal graph network by a causal DAG, while we refer a Bayesian graph network by a DAG. The terminologies in a causal DAG model are as follows [4]:

- We assign a random variable X_i to each vertex in $V \in \{x, y, z_1, z_2, z_3\}$ (X_x, X_y, etc.), and the edges E denote relationships between pairs of random variables.
- If $z_1 \to x$ then z_1 is a *parent* of x, and x is a *child* of z_1.
- If there is a *directed path* from z_1 to y then z_1 is an *ancestor* of y, and y is a *descendant* of z_1; each vertex is also an *ancestor* and *descendant* itself.
- The sets of *parents*, *children*, *ancestor*, and *descendants* of y in our graph G are denoted by $pa(y, G)$, $ch(y, G)$, $an(y, G)$, and $desc(y, G)$.

If we consider a subset $S \in \{X_{z_1}, X_{z_3}, X_x, X_y\}$ then we can write the dependence model according to a first-order Markovian model:

$$X_y \perp\!\!\!\perp X_{z_1} | \{X_x, X_{z_3}\}, \tag{1}$$

which implies that the future is independent of the past given the present (see Fig. 1 [Left]). Moreover, a set of variables $X_{pa(y)}$ is said to be *Markovian parents* of X_y if it is a minimal subset of $\{X_{z_1}, X_{z_3}, X_x, X_y\}$ such that $p(X_y|X_{z_1}, X_{z_3}, X_x) = p(X_y|X_{pa(y)})$. We can re-express the equation (1) given a subset $S_1 = \{X_y\} \in S$ by

$$X_{S_1} \perp\!\!\!\perp X_{nondesc(S_1)\backslash pa(S_1)} | X_{pa(S_1)}, \tag{2}$$

where $X_{nondesc(S_1)\backslash pa(S_1)} = \{X_{z_1}, X_{z_3}, X_x\} \backslash \{X_{z_3}, X_x\} = \{X_{z_1}\}$, and $X_{pa(S_1)} = \{X_x, X_{z_3}\}$.

Furthermore, a DAG model, that is a Bayesian graph network, is defined by a pair (G, P)

$$p(X_{z_1}, X_{z_3}, X_x, X_y) = p(X_{z_1}|X_{pa(z_1)}) p(X_{z_3}|X_{pa(z_3)}) p(X_x|X_{pa(x)}) p(X_y|X_{pa(y)}), \tag{3}$$

where p is factorized according to the DAG G, and p is a density function, and in general, for any graph with a subset $S = X_v$,

$$p(S) = \Pi_{j \in v} p(X_j|X_{pa(j)}). \tag{4}$$

In our case, a subset X_v was $\{X_{z_1}, X_{z_3}, X_x, X_y\}$, namely, the index v was $\{z_1, z_3, x, y\}$.

We now consider that G represents real-world measurement data points, and its edges represent the causal relations. If we intervene this DAG model (G, p) by setting $X_x = X_{x'}$ then we have

$$p(X_{z_1}, X_{z_3}, X_x, X_y | do(X_x = X_{x'})) = p(X_{z_1}|X_{pa(z_1)}) p(X_{z_3}|X_{pa(z_3)}) I(X_x = X_{x'}) p(X_y|X_{pa(y)}), \tag{5}$$

which we graphically illustrated in Fig. 1 [Right] [7]. Moreover, we intervened a data point X_x by setting it to the value $X_{x'}$. We must note that, here, $X_{x'}$ is a number while X_x takes a random value. As a result, a DAG G becomes a truncated graph $\bar{G}$ as illustrated in Fig. 1; the intervention is denoted by the operator $do(X_x = X_{x'})$.

Hence, a causal DAG model (i.e. a causal Bayesian model) for any graph can be expressed by

$$p(X_v|do(X_u = X_{u'})) = \Pi_{j \in v \backslash u} p(X_j|X_{pa(j)}) I(X_u = X_{u'}), \tag{6}$$

where v notes a set of nodes, and u is a single node index or an intervened data point.

N	Number of causal DAGs
0	1
1	1
2	3
3	25
4	543
5	29281
6	3781503
7	1138779265
8	783702329343
9	1213442454842881
10	4175098976430598143
11	31603459396418917607425
12	521939651343829405020504063
13	18676600744432035186664816926721
14	1439428141044398334941790719839535103

Fig. 2. A number of causal DAGs given N data points.

3. OUR TOY DATA

Remote sensing datasets are obtained by sensors on satellites and airbornes, and different benchmark datasets are currently available [8], [9]. To investigate a causal DAG model for remote sensing, we use three scene images, *cloud*, *water*, and *agriculture*. We emphasize again that the causal structure of scene images is known to avoid discovering causal relations [6, 10, 11] since discovering a causal relation is a hard problem; for example, If we consider $N = 5$ data points with unknown causal relations then there are $29,281$ possible causal DAGs [12] (see Fig. 2).

Furthermore, we characterize each scene image by a random variable X_i (i.e. entropy) shown in Fig. 3 [Top] since scenes can change over time due to the external effects. Moreover, we consider the nodes z_1 and z_2 as unknown confounders, and we assume that *clouds* are the source of information instead of the source of issues [13]. Hence, we study total causal effects on *agriculture* by intervening the *water* value X_x and the *cloud* value X_{z_3} shown in Fig. 3 [Bottom], respectively. Namely, we obtain information on *agriculture* when measuring that either there is a change in *water* quantity, or there is *cloud* coverage over a certain period of time.

4. METHODOLOGY: LINEAR STRUCTURAL EQUATION MODEL

We represent a causal DAG by using a linear Structural Equation Model (a linear SEM) which refers to a linear model L such that a node X_j is an effect of its parents with an additive noise ϵ_j,

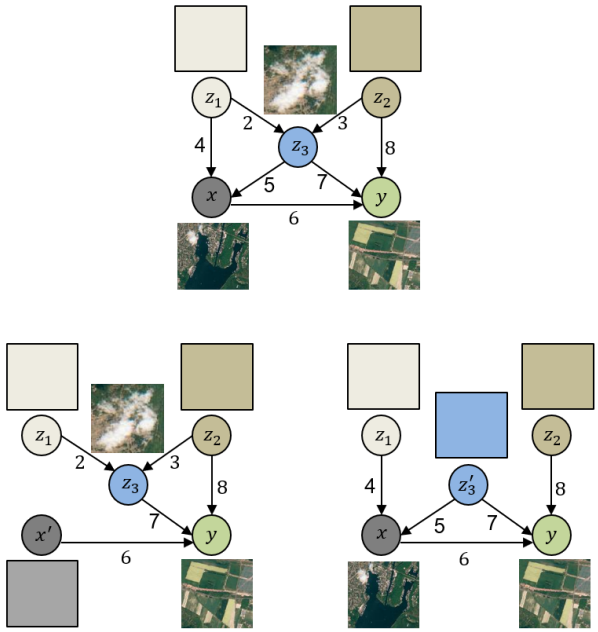

Fig. 3. The considered causal DAG model for remote sensing: [Top] The original DAG model G, [Bottom] The intervened DAG models G' when $do(X_x = X_{x'})$ and $do(X_{z_3} = X_{z_3'})$, respectively.

$$X_j = L(X_{pa(j)}, \epsilon_j). \tag{7}$$

First remark: If we intervene $do(X_j = X_{j'})$ then we simply replace $L(X_{pa(j)}, \epsilon_j)$ by $X_{j'}$, and a causal DAG becomes a truncated causal DAG:

$$X_j = X_{j'}. \tag{8}$$

Second remark: The total causal effects of X_i on X_j can be found by multiplying the weights of the directed edges X_i to X_j in each path, and then summing up the multiplied weights for each path.

5. LINEAR STRUCTURAL EQUATION MODEL FOR OUR TOY DATASET

we express a linear SEM for our toy dataset by

$$\begin{aligned} X_{z_1} &= \epsilon_1(0,1), \\ X_{z_2} &= \epsilon_2(0,1), \\ X_{z_3} &= 2X_{z_1} + 3X_{z_2} + \epsilon_3(0,1), \\ X_x &= 4X_{z_1} + 5X_{z_3} + \epsilon_4(0,1), \\ X_y &= 6X_x + 7X_{z_3} + 8X_{z_2} + \epsilon_5(0,1), \end{aligned} \tag{9}$$

where $\epsilon_k(0,1), \quad k = 1, 2, \ldots, 5$, is a Gaussian random variable with $\mu = 0, \sigma = 1$ (see Fig. 3 [Top]) [4], and we simulate our experiment by using $n = 100$ data points. First, we intervene our *water* quantity by setting it to a constant value zeros $X_{x'} = \mathbf{0}$ and ones $X_{x'} = \mathbf{1}$. We assume that there is no change in *water* quantity $do(X_x = \mathbf{0})$ and a constant change $do(X_x = \mathbf{1})$ over a given time frame (see Fig. 3 [Bottom] left).

$$\begin{aligned} X_{z_1} &= \epsilon_1(0,1), \\ X_{z_2} &= \epsilon_2(0,1), \\ X_{z_3} &= 2X_{z_1} + 3X_{z_2} + \epsilon_3(0,1), \\ X_x &= X_{x'} = \mathbf{0}, \\ X_y &= \mathbf{0} + 7X_{z_3} + 8X_{z_2} + \epsilon_5(0,1) \end{aligned} \tag{10}$$

and

$$\begin{aligned} X_{z_1} &= \epsilon_1(0,1), \\ X_{z_2} &= \epsilon_2(0,1), \\ X_{z_3} &= 2X_{z_1} + 3X_{z_2} + \epsilon_3(0,1), \\ X_x &= X_{x'} = \mathbf{1}, \\ X_y &= \mathbf{1} + 7X_{z_3} + 8X_{z_2} + \epsilon_5(0,1). \end{aligned} \tag{11}$$

The total causal effect on *agriculture* is then by the intervention $do(X_x = X_{x'})$:

$$\mathbb{E}[X_y | X_{x'} = \mathbf{1}] - \mathbb{E}[X_y | X_{x'} = \mathbf{0}] = 6 \neq 0. \tag{12}$$

Secondly, we intervene our *cloud* coverage to obtain causal information on *agriculture*, and in particular, we set $X_{z_3} = \mathbf{0}$ If there is no *cloud* coverage over a certain period of time, and we assign $X_{z_3} = \mathbf{1}$ if there is *cloud* coverage over a certain period of time. Moreover, we compute a total causal effect on *agriculture* and attempt to model *agriculture* independent of *cloud* coverage. Hence, the SEM becomes

$$\begin{aligned} X_{z_1} &= \epsilon_1(0,1), \\ X_{z_2} &= \epsilon_2(0,1), \\ X_{z_3} &= X_{z_3'} = \mathbf{0}, \\ X_x &= 4X_{z_1} + 5\mathbf{0} + \epsilon_4(0,1), \\ X_y &= 6X_x + 7\mathbf{0} + 8X_{z_2} + \epsilon_5(0,1) \end{aligned} \tag{13}$$

and

$$\begin{aligned} X_{z_1} &= \epsilon_1(0,1), \\ X_{z_2} &= \epsilon_2(0,1), \\ X_{z_3} &= X_{z_3'} = \mathbf{1}, \\ X_x &= 4X_{z_1} + 5\mathbf{1} + \epsilon_4(0,1), \\ X_y &= 6X_x + 7\mathbf{1} + 8X_{z_2} + \epsilon_5(0,1). \end{aligned} \tag{14}$$

The total causal effect on *agriculture* is then by the intervention $do(X_{z_3} = X_{z'_3})$:

$$\mathbb{E}[X_{z_3}|X_{z'_3} = \mathbf{1}] - \mathbb{E}[X_{z_3}|X_{z'_3} = \mathbf{0}] = 37 \neq 0. \quad (15)$$

These total causal effects (12) and (15) are equal to zero if there is no a cause-effect relation. Our total causal values are not equal to zero in both of our intervention experiments presented in equations (12) and (15). The results of our intervention experiments demonstrate that we can model and predict *agricultural* lands by using information on *water* quantity and *cloud* coverage when we know their causal structure. In particular, there is the advantage that we can obtain *water* quantity data from the *Sentinel-1* satellite and *cloud* coverage data from the *Sentinel-2* satellite.

In our causal model, the *cloud* coverage becomes a very important information element. More importantly, we see that the *cloud* coverage can explain strongly the *agriculture* land change than the *water* quantity, and it is evidently proved by Eqs. (12) and (15): A higher cause-effect relation (i.e. a higher expected value) between the *cloud* coverage and the *agriculture* land than the *water* quantity and the *agriculture* land.

6. CONCLUSION

In this paper, we proposed to encode each patch of our toy dataset in a random variable X_i of the SEM, and we investigated the SEM by intervening on *water* quantity and *cloud* coverage when we know the causal structure of our toy dataset. We discovered that the cause-effect relation (i.e. expected value) between *cloud* coverage and the *agriculture* land is higher than the *water* quantity and the *agriculture* land. As an ongoing and future work, we integrate our causal model in Deep Learning and Machine learning methods, and we design an algorithm for discovering a causal relation of data points with a unknown causal relation [14].

7. REFERENCES

[1] M. Datcu, K. Seidel, and M. Walessa, "Spatial information retrieval from remote-sensing images. i. information theoretical perspective," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1431–1445, 1998.

[2] M. Schroder, H. Rehrauer, K. Seidel, and M. Datcu, "Spatial information retrieval from remote-sensing images. ii. gibbs-markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1446–1455, 1998.

[3] Judea Pearl, "Causal inference in statistics: An overview," *Statistics Surveys*, vol. 3, no. none, pp. 96 – 146, 2009.

[4] J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference - Foundations and Learning Algorithms*, Adaptive Computation and Machine Learning Series. The MIT Press, Cambridge, MA, USA, 2017.

[5] Adrián Pérez-Suay and Gustau Camps-Valls, "Causal inference in geoscience and remote sensing from observational data," 2020.

[6] Jakob Runge, Xavier-Andoni Tibau, Matthias Bruhns, Jordi Muñoz Marí, and Gustau Camps-Valls, "The causality for climate competition," in *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, Hugo Jair Escalante and Raia Hadsell, Eds. 08–14 Dec 2020, vol. 123 of *Proceedings of Machine Learning Research*, pp. 110–120, PMLR.

[7] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio, "Towards causal representation learning," *CoRR*, vol. abs/2102.11107, 2021.

[8] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang, "Dota: A large-scale dataset for object detection in aerial images," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[9] Gencer Sumbul, Marcela Charfuelan, Begüm Demir, and Volker Markl, "Bigearthnet: A large-scale benchmark archive for remote sensing image understanding," *CoRR*, vol. abs/1902.06148, 2019.

[10] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio, "Towards causal representation learning," 2021.

[11] Johann Brehmer, Pim de Haan, Phillip Lippe, and Taco Cohen, "Weakly supervised causal representation learning," 2022.

[12] "A Number of DAGs with N labeled nodes: http://oeis.org/a003024," 2022.

[13] Andrea Meraner, Patrick Ebel, Xiao Xiang Zhu, and Michael Schmitt, "Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 333–346, 2020.

[14] Soronzonbold Otgonbaatar and Mihai Datcu, "A quantum annealer for subset feature selection and the classification of hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7057–7065, 2021.