Uncovering Causal Relationships in Co-location Patterns: Approximating Direct Causes through Granger Causality Mining

Hong Su
hsuag@connect.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

Raymond Chi-Wing Wong
raywong@cse.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

ABSTRACT
Mining causal relationships within co-location patterns is a crucial aspect of knowledge discovery, with broad applications spanning across ecosystems, praxiology, and epidemiology. However, existing approaches to solving this problem still have limitations. Some causality models assume rigid definitions of causation that may not generalize or do not fully leverage spatial-temporal information. Moreover, how to differentiate direct causes and indirect causes is also challenging. To address them, this paper proposes a generalized model for mining causality from co-location patterns. Our approach, rooted at Granger causality, integrates an algorithm that approximates direct causes from Granger causes by leveraging unique linear causal information. We conducted extensive experiments on real-world datasets to evaluate the effectiveness of our method and compared it with two baselines.

CCS CONCEPTS
• Information systems → Spatial-temporal systems.

ACM Reference Format:

1 INTRODUCTION
In the age of big data, the abundance of location-based data obtained from GPS-equipped electronic devices presents an opportunity for mining rules of interest. Spatial-temporal datasets, encompassing time, location, and other relevant attributes, hold significant value across ecosystems, praxiology, and epidemiology. However, existing causality models suffer from several challenges that may not generalize or do not fully leverage spatial-temporal information. Moreover, how to differentiate direct causes and indirect causes is also challenging. To address them, this paper proposes a generalized model for mining causality from co-location patterns. Our approach, rooted at Granger causality, integrates an algorithm that approximates direct causes from Granger causes by leveraging unique linear causal information. We conducted extensive experiments on real-world datasets to evaluate the effectiveness of our method and compared it with two baselines.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGSPATIAL ’23, November 13-16, 2023, Hamburg, Germany
© 2023 Association for Computing Machinery.
ACM ISBN 978-8-8007-0168-9/23/11...
https://doi.org/10.1145/3589132.3628373

2 METHODOLOGY
Firstly, we introduce some preliminaries.

Co-location Patterns [1]. If some types of objects form a co-location pattern, these types of objects often co-occur with each other within a specific geographic area.

Partial Correlation. A partial correlation quantifies conditional independence, which measures correlation between two variables after removing the linear influence of other variables. Formally, the partial correlation between two real variables X and Y conditioned on another real variable Z, denoted as $\rho_{XY|Z}$, is computed...
where 1) variable used in Granger causality test of temporal precedence, we further filter the primary Granger causes (selected Granger causes for Ours and DAG causes for NO-TEARS).

Granger Causality [3]. Given two time series \( \{x_t\} = \{x_1, x_2, \cdots \} \) and \( \{y_t\} = \{y_1, y_2, \cdots \} \), where \( x_t \) and \( y_t \) mean the data at timestamp \( t \) for \( t \in \{1, 2, 3, \cdots \} \), the Granger causality test is that, we add the lagged values of \( \{x_t\} \) to \( \{y_t\} \), i.e., from the reduced model \( y_t - a_0 + a_1 y_{t-1} + \cdots + a_L y_{t-L} + \epsilon_t \) to the full model \( y_t = a_0 + a_1 y_{t-1} + \cdots + a_L y_{t-L} + b_1 x_{t-1} + \cdots + b_L x_{t-L} + \epsilon_t \), where \( \epsilon_t \) is Gaussian white noise (a real number). For co-location patterns \( A \) (with data \( \{x_t\} \)) and \( B \) (with data \( \{y_t\} \)), if augmenting \( A \)’s history helps predict \( B \)’s future (i.e., the aforementioned full model outperforms the reduced model significantly), we say that \( A \) is a Granger cause of \( B \) and \( A \) contains linear causal information about \( B \). Besides, the causes in our paper are defined as Granger causes.

Co-location Increment. Let \( \{N'_t\} \) be the time series of the numbers of a co-location pattern \( c \) at different time \( t \). We define the variable used in Granger causality test of \( c \) to be \( \Delta N'_t = N'_t - N'_{t-1} \) (i.e., \( x_t, y_t \) to be \( \Delta N'_t \) of some patterns).

Unique Linear Causal Information. We illustrate it by the example \( A \rightarrow B \rightarrow C \). Granger causality test is a linear test and partial correlation \( \rho_{\Delta N^A \Delta N^C | \Delta N^B} \) involving removes linear information of direct causes (\( B \)) on indirect causes (\( A \)) via \( \Delta N^A - L(\Delta N^A | \Delta N^B) \).

Let \( B \)’s linear causal information be \( \{\phi \cup \xi\} \) and \( A \)’s be \( \{\phi' \cup \delta\} \), where 1) \( \phi \) is linear causal information about \( C \), 2) \( \phi' \subset \phi \) and 3) \( \xi, \delta \) are linear causal information about other patterns disjoint with \( \phi \), \( \phi' \subset \phi \) because \( A \)’s linear causal information about \( C \) is conveyed by \( B \) and there exists some information loss during the conveyance. If \( A \) removes \( B \)’s linear information including linear causal information (i.e., \( \phi, \xi \)), the residuals \( \Delta N^A - L(\Delta N^A | \Delta N^B) \) contain linear causal information no more than \( \delta \), i.e., \( A \) has insignificant unique linear causal information about \( C \) and \( \rho_{\Delta N^A \Delta N^C | \Delta N^B} \) is close to zero.

We then introduce our algorithm as follows, which contains two parts. First, we identify the Granger causes for each prevalent co-location pattern. However, since Granger causes only ensure temporal precedence, we further filter the primary Granger causes to eliminate some indirect causes. In the second component, we propose a select-and-prune algorithm for filtering causes, based on unique linear causal information. For a co-location pattern \( c \), we assume that its direct cause, say \( b \), can predict it most accurately (e.g., based on MSE). Then, we compute the partial correlation between \( \Delta N_c \) of each of the causes among \( c \)’s remaining Granger causes and \( c \) conditioned on that of \( b \) and prune those causes with a small computed partial correlation. For example, consider \( A \rightarrow B \rightarrow C \). Here, as described above, \( C \) is the co-location pattern (\( c \)) and \( B \) is the direct cause (\( b \)). We first mine \( A \) and \( B \) as the candidate Granger causes of \( C \). Then, for selection, we first select \( B \). Then, we compute \( \rho_{\Delta N^A \Delta N^C | \Delta N^B} \). A will be pruned because \( \rho_{\Delta N^A \Delta N^C | \Delta N^B} \) is small.

### 3 EXPERIMENT RESULTS

We used the Foursquare NYC Check-in\textsuperscript{3} dataset (227,428 records, 1,083 users, 251 venues, 10 months). We compared our models with CRDA and NO-TEARS. We treated our results as ground truths and some statistics are in Table 1. (1) CRDA only considers 3 consecutive timestamps originally. We adapted it by randomly choosing 3 consecutive timestamps from the total 288 timestamps. CRDA identifies only a few causal rules with low recall and precision and had a lengthy computation time due to its rigid causation definition and the need to compute numerous pair-wise distances. We did not compare our model with CRDA on all timestamps due to its low efficiency. (2) NO-TEARS also has low recall and precision because the DAG causes (ignoring time) have different physical meanings. Figure 1 highlights the absence of cyclic causality chains and the presence of isolated objects like Hotel. In contrast, our results seem more reasonable, as they show expected relationships, such as School being related to General Education Place.

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of rules</th>
<th>Recall</th>
<th>Precision</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRDA</td>
<td>9.8</td>
<td>4.6%</td>
<td>20.1%</td>
<td>10237.6</td>
</tr>
<tr>
<td>NO-TEARS</td>
<td>31</td>
<td>12.8%</td>
<td>16.1%</td>
<td>31.3</td>
</tr>
<tr>
<td>Ours</td>
<td>39</td>
<td>-</td>
<td>-</td>
<td>1.6</td>
</tr>
</tbody>
</table>

### 4 CONCLUSION

We propose a generalized problem definition and framework for causality mining in co-location patterns by integrating spatial information into Granger causality. Furthermore, we efficiently filters out some indirect causes based on unique linear causal information. Our experimental results demonstrate promising results in generating reasonable causality chains. Future work can be better quantitatively evaluating the effectiveness of causality mining.

### REFERENCES


