Research on the Discovery of Timing Causal Structure in Epidemic Prevention and Control

Tengjiao Mao, Chunxiao Cai, Yue Lu*, Ruihua Wang and Wei Liu
Center for Strategic Assessment and Consulting, AMS, Beijing 100097, China

* *Corresponding author*

Keywords: Epidemic Prevention and Control, Timing Causation, Dimensionality Reduction, Convergence.

Abstract: Decision-makers need to timely and accurately identify departments with unfavorable handling based on the effect of epidemic prevention and control, which requires the construction of a timing causal structure between departments. In consideration of a large number of departments and influencing factors involved, traditional algorithms are more costly to construct causal structures. In this paper, the departments involved in epidemic prevention and control and related factors are analyzed. A causality analysis framework based on Bayesian networks is proposed. The dimensionality of data is reduced based on time-varying characteristics. Bayesian network structure learning algorithms are used to build a structural model based on timing causality. The results of the simulation case show that the method takes the advantage of fast convergence and accurate causality.

1 INTRODUCTION

Inferring causal relationships between things is a hot topic in the study of data relationships. Causal inference methods, as the main means of causality research, have been widely used in the fields of policy evaluation, fault diagnosis, biomedicine, etc. For example, researchers have applied causal inference models in disease diagnosis, biological network inference, and drug efficacy analysis in the biomedical field (Sesia, 2020; Shen, 2020; Zhou, 2010; Dong, 2014; Cai, 2013; Liu, 2014). In the field of communication and industry, some scholars have applied Bayesian networks to fault diagnosis and performance optimization studies of networks (Hao, 2016; Trave-Massuyes, 1997; Hu, 2013). In the field of social networks, researchers have tried to use causal discovery models to study the causal relationships of user behaviors (Ver Steeg, 2012; Duan, 2013; Sun, 2014; Sun, 2015).

Exploring the causal relationship between various departments in epidemic prevention and control is also needed. Especially in the case of poor prevention and control, a higher degree of accuracy and timeliness is required to trace the problem and allocate responsibility, so that the relevant departments can be targeted to rectify and effectively curb the spread of the epidemic on time. Therefore, it is relevant to research the discovery method of the causal structure of departments in epidemic prevention and control based on timing multidimensional data.

The widely used causal inference methods are mainly divided into Rubin's potential outcome model and Pearl's SCM (structure causal model). Causal structure discovery methods mainly rely on causal structure learning models, which are mainly classified as conditional independence constraint-based methods (Pearl, 2009; Le, 2016; Cai, 2011; Tu, 2019) and scoring-based methods (Ramsey, 2017; Huang, 2018; Zheng, 2018). Conditional independence constraint-based methods learn the causal structure by judging the independence and conditional independence information between nodes, and typical algorithms are the PC (Peter-Clark) algorithm and IC (Inductive Causation) algorithm. Scoring-based methods discover causal structures by scoring them based on observed data, typical methods are GES (Greedy equivalence search) and FGES (fast greedy equivalence search).

However, there are many departments and emergencies involved in epidemic prevention and control, and the factors that affect the handling capabilities of departments result in large changes in the timing epidemic prevention results. To accurately learn the causal structure among departments, a great amount.

Mao, T., Cai, C., Lu, Y., Wang, R. and Liu, W.
Research on the Discovery of Timing Causal Structure in Epidemic Prevention and Control.
DOI: 10.5220/0011737900003607
In *Proceedings of the 1st International Conference on Public Management, Digital Economy and Internet Technology (ICPDI 2022)*, pages 386-392
ISBN: 978-989-758-620-0

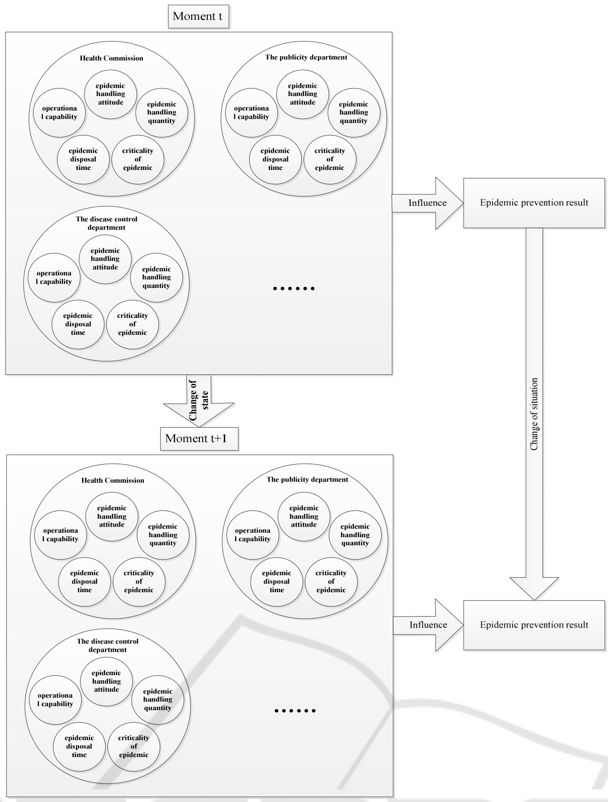


Figure 1: Diagram of epidemic prevention and control based on timing changes.

of multi-dimensional timing sample data is required. The learning cost is too high if traditional algorithms are used directly. To solve the above problems, the key elements in epidemic prevention and control are analyzed and summarized. The dimensionality of the data is reduced based on the timing change characteristics. Timing causal structures are constructed through scoring-based learning algorithms. Finally, simulated data are utilized to verify the effectiveness of the algorithm. The simulation results indicate that the proposed algorithm can effectively reduce the dimensionality of data, which results in faster convergence speed and lower learning costs

2 CONSTRUCTION OF MULTI-DIMENSIONAL TIMING NODE CAUSAL STRUCTURE DISCOVERY MODEL

2.1 Analysis of Key Departments and Influencing Factors

A city's emergency plan for epidemic prevention and control is taken as an example in the paper (Shanghai municipal health commission, 2020). The departments mainly involved in epidemic prevention and control include the Health Commission, the publicity departments, the grassroots communities, hospitals, and the disease control departments. Their main functions in epidemic prevention and control are summarized as follows.

The Health Commission is responsible for formulating disease prevention plans and immunization plans, and implementing intervention measures for public health problems which endanger people's health.

The publicity department is responsible for publicizing epidemic prevention policies, monitoring and controlling online public opinion, etc.

The hospital is responsible for treating patients and assisting in disease testing and virus elimination.

The disease control department is responsible for epidemiological investigation of cases, case closure and virus elimination efforts.

The grassroots community is responsible for monitoring the health of people in the community, regularly organizing disease screenings, and conducting preliminary disposition, reporting, and prevention and control once cases are detected.

The departments above all play an important role in epidemic prevention and control. When emergencies come, each department needs to handle the situation individually or collaboratively. There are also many factors that affect the handling ability of departments, including operational capability, epidemic handling attitude, epidemic handling quantity, epidemic disposal time, and severity of the epidemic. With the change of time and the influence of various factors, the handling capacity of each department will not always remain the same. Changes in handling capacity will affect the result of epidemic prevention and control, as shown in Fig.1.

Because of the great number of departments and factors involved and the timing change, it is necessary to construct a causal structure to timely discover the influence relationship among departments.

2.2 Algorithm Model Construction

In Pearl's structural causal model (Pearl, 2009), a DAG (directed acyclic graph) is used to construct the causal structure. In this paper, departments and prevention results are used as variable nodes.

Take the simplest case as an example. The handling capacity of departments is influenced by the above five factors., and it is assumed that each factor has only two states which are denoted as 0 and 1 respectively at moment t. The state of the node represented by each department at that moment has 2^5 possible values which are noted as binary from 00000 to 11111. As an outcome node, the epidemic prevention result is usually measured by the number of infected people at moment t.

Each node takes 5 factors into consideration, resulting in an excessive number of possible values. It will greatly reduce the efficiency of the structure learning algorithm if the causal structure is learned in this way. It is necessary to improve the model and reduce the dimensionality of the data. The specific dimensionality reduction method is as follows.

In the original data, a 5-bit binary number is taken to represent the value of the node state. The state of each node at different times is counted, and the state of nodes at two adjacent moments is compared. If the value of node state remains unchanged, it is recorded as 0, and if the value of node state changes, it is recorded as 1. Therefore, 1-bit binary number can be used to represent the state changes of department nodes in the adjacent periods, so as to realize the dimension reduction of the number of department node values, as is shown in Fig.2. Considering the changes in epidemic prevention results, when the number of infected people increases, it can be considered that the prevention results have become worse (recorded as -1); when the number of infected people falls, the prevention results can be considered better (recorded as 1); when the number of infected people remains unchanged, the prevention results can be considered unchanged (recorded as 0).

As time advances, multiple sets of data on node state changes will be formed. Finally, according to the dimensionality-reduced data, the corresponding causal graph is derived with a score-based Bayesian structure learning algorithm.

3 SIMULATION VERIFICATION

To verify the validity of the model in this paper, the causal structure between departments is preset in the context of epidemic prevention and control in a city, and the timing change data is generated according to this structure. Based on the generated data, the traditional learning algorithm and proposed algorithm are utilized to learn the causal structure separately, and the learning results and costs are compared and analyzed to verify the validity of the algorithm.

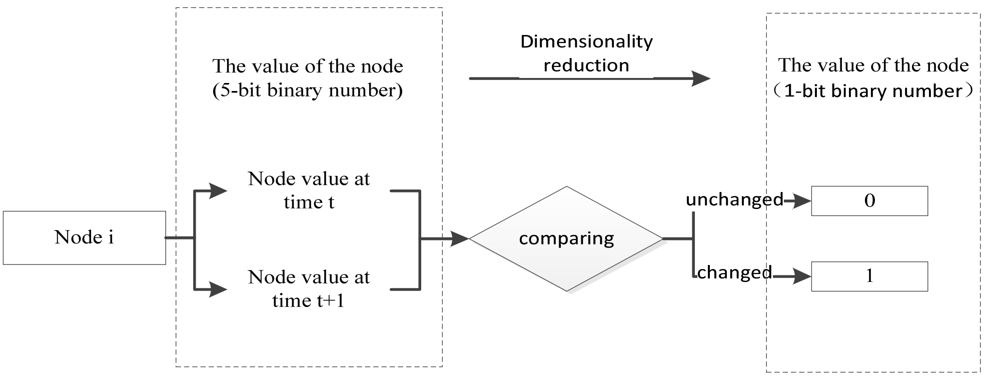


Figure 2: Diagram of node value dimensionality reduction.

The simulation flow is shown in Fig.3.

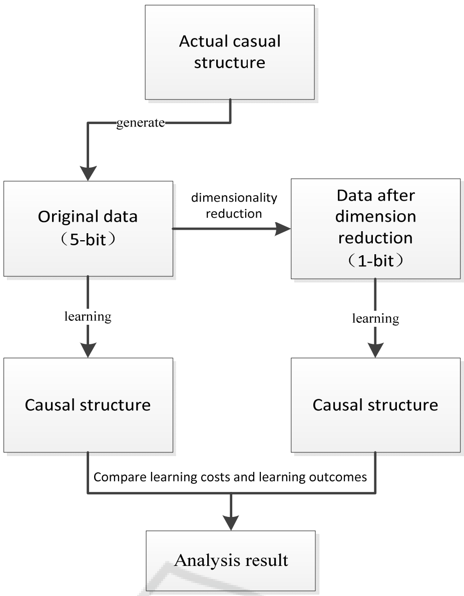


Figure 3: The diagram of simulation flow.

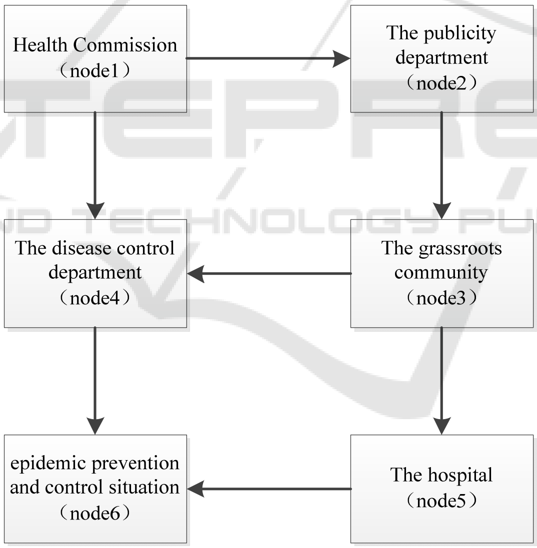


Figure 4: summarized causal structure.

3.1 Data Generation and Dimensionality Reduction

According to the contents of the emergency management plan and work measures for municipal epidemic prevention and control (Shanghai municipal health commission, 2020), the influence relationship between departments and the epidemic prevention result is qualitatively summarized as shown in Fig.4.

Different moments are divided evenly in enough time. At each moment, the state of each node is generated with a certain conditional probability according to the influence relationship in Fig.4. In this way, the state value (5-bit) of each node is generated as sample data. Assuming that the sample data at each moment belongs to independent and identical distribution, multiple sets of simulated data are generated in this way. Traditional algorithms perform causal structure learning based on this data.

The dimensionality of the simulated data is reduced with the proposed algorithm. And multiple sets of node state change data (1-bit) for different time periods are formed. The proposed algorithm learns the causal structure based on the data.

3.2 Criteria For Comparison of Results

In order to measure the difference between the proposed algorithm which learns after dimensionality reduction and the traditional algorithm, the results are compared and analyzed in terms of learning effect and learning cost.

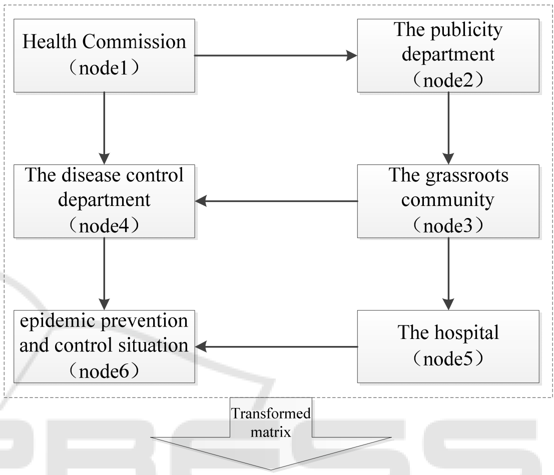
The learning cost is measured by the size of samples and the time used for learning. The sample size plays an important role in the result of causal structure learning, and the accurate structure always cannot be obtained with little data. Theoretically, the more sample data is learned, the more accurate the learned causal structure will be. When a certain amount of sample data is exceeded, the learned causal structure will not change. It can be considered that the algorithm has converged and learning has completed. The size of samples and the learning time is used as the learning cost when the algorithm converges.

When the algorithm has converged, the effect of learning is measured by the similarity between the learned causal structure and the actual causal structure. If the structure in Fig.4 is assumed to be the actual causal structure, comparing the causal structure at the time of algorithm convergence with the causal structure in Fig.4, it can be concluded that a higher degree of similarity represents a more accurate learned causal structure. In order to quantify the degree of similarity between two causal structures, the concept of Structural Relevance is defined.

Structural Relevance $\boldsymbol{R}$ (Adler, 2010): Structural Relevance quantitatively characterizes the degree of similarity between two causal structures. The causal structure can be transformed into the form of a matrix. If there are $\boldsymbol{n}$ nodes in the causal structure, it can be represented by a $\boldsymbol{n}\times\boldsymbol{n}$ matrix. When node $\boldsymbol{i}$ in the causal structure has an arrow pointing to the node $\boldsymbol{j}$ (node $\boldsymbol{i}$ is the cause of node $\boldsymbol{j}$), the value $(\boldsymbol{i},\boldsymbol{j})$ in the corresponding matrix is 1, otherwise, it is 0. Taking the actual causal structure as an example, the schematic diagram of its causal structure transformed into a matrix is shown in Fig.5. Then, the degree of similarity of two causal structures can be expressed in terms of the similarity of two matrices. The matrix transformed by the actual causal structure is denoted as $\boldsymbol{A}$, and the matrix transformed by the matrix obtained from learning is denoted as $\boldsymbol{B}$. The specific formula for the structural correlation $\boldsymbol{R}$ is as follows.

$$R=\frac{\sum_n\sum_n\left(A_{nn}-\bar{A}\right)\left(B_{nn}-\bar{B}\right)}{\sqrt{\left(\sum_n\sum_n\left(A_{nn}-\bar{A}\right)^2\right)\left(\sum_n\sum_n\left(B_{nn}-\bar{B}\right)^2\right)}}$$

$\bar{\boldsymbol{A}}$ (or $\bar{\boldsymbol{B}}$) is equivalent to averaging the value of all elements in $\boldsymbol{A}$ (or $\boldsymbol{B}$), and $\boldsymbol{n}$ is the matrix rank. The closer the value $\boldsymbol{R}$ is to 1, the more similar the two structures are meant to be. When $\boldsymbol{R}=1$, it indicates that the actual causal structure is learned.



$$\begin{array}{c} \\ \text{node1}\\ \text{node2}\\ \text{node3}\\ \text{node4}\\ \text{node5}\\ \text{node6}\end{array}\begin{array}{c}\begin{array}{cccccc}\text{node1}&\text{node2}&\text{node3}&\text{node4}&\text{node5}&\text{node6}\end{array}\\ \begin{bmatrix}0&1&0&1&0&0\\0&0&1&0&0&0\\0&0&0&1&1&0\\0&0&0&0&0&1\\0&0&0&0&0&1\\0&0&0&0&0&0\end{bmatrix}\end{array}$$

Figure 5: Schematic diagram of the actual causal structure transformation matrix.

3.3 Results Display and Analysis

With the traditional algorithm and the proposed algorithm in this paper, the correct causal structure obtained after the final algorithm convergence is shown in Fig.6.

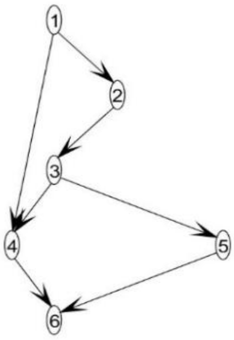


Figure 6:Learned causal structure.

The convergence of the algorithm is shown in Fig.7.

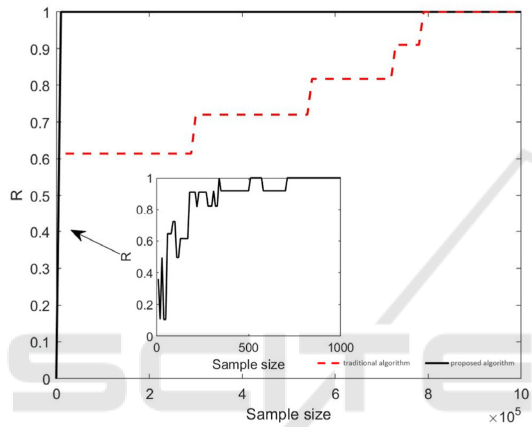


Figure 7: The diagram of algorithm convergence.

Both methods learned the actual causal structure from Fig.6. But there are significant differences in learning cost. According to Fig.7, 800000 samples are required for the traditional algorithm to learn the actual causal structure. While for the proposed algorithm, only 800 samples are needed to learn the actual causal structure. The size of samples learned by this algorithm is reduced by 1000 times. From the perspective of learning time, the convergence time of the traditional algorithm is 11250s, while the convergence time of the proposed algorithm is 10s. The time cost is reduced by 1125 times.

It can be seen that the proposed algorithm learns an accurate causal structure based on the simulated data, and greatly outperforms the traditional algorithm in terms of time cost and the size of samples, which verifies the advantages and effectiveness of the proposed algorithm. When using real data, this method is equally applicable and advantageous.

4 CONCLUSION

In order to construct an accurate causal structure and reduce the cost of learning, the dimension of timing data is reduced based on the changing features and the causal structure is constructed through the structural learning algorithms in this paper. The proposed algorithm and the traditional algorithm are compared in learning effect and cost with an example. The proposed algorithm takes the advantage of fast convergence and accurate causality. However, the size of samples required by the proposed algorithm is still considerable in actual needs, so the size of samples required for learning needs to be further reduced in the next research to achieve better results.

REFERENCES

Adler J, Parmryd I. Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient[J]. Cytometry Part A, 2010, 77a(8):733-742.

Cai R, Zhang Z, Hao Z. BASSUM: A Bayesian semi-supervised method for classification feature selection[J]. Pattern Recognition, 2011, 44(4): 811-820.

Cai R, Zhang Z, Hao Z. Causal gene identification using combinatorial V-structure search[J]. Neural Networks, 2013, 43: 63-71.

Dong G, Wang Y, Zhang Q, et al. The methodology of Dynamic Uncertain Causality Graph for intelligent diagnosis of vertigo[J]. Computer Methods and Programs in Biomedicine, 2014, 113(1):162-174.

Duan P, Yang F, Chen T, et al. Direct causality detection via the transfer entropy approach[J]. IEEE Transactions on Control Systems Technology, 2013, 21(6): 2052-2066.

Hao Z, Chen W, Cai R, et al. Performance optimization of wireless network based on canonical causal inference algorithm [J]. Journal of Computer Applications, 2016, 36(8): 2114-2120.

Hu Y, Zhang X, Ngai E W T, et al. Software project risk analysis using Bayesian networks with causality constraints[J]. Decision Support Systems, 2013, 56: 439-449.

Huang B, Zhang K, Lin Y, et al. Generalized score functions for causal discovery[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery&Data Mining, 2018:1551-1560.

Le T, Hoang T, Li J, et al. A fast PC algorithm for high dimensional causal discovery with multi-core PCs[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2016, 16(5): 1483-1495.

Liu M, Cai R, Hu Y, et al. Determining molecular predictors of adverse drug reactions with causality analysis based on structure learning[J]. Journal of the American Medical Informatics Association, 2014, 21(2): 245-251.

Pearl J. Causality: models, reasoning and inference[M]. Cambridge University Press, 2009: 107-173.

Ramsey J, Glymour M, Sanchez-Romero R, et al. A million variables and more: the Fast Greedy Equivalence Search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images[J]. International Journal of Data Science and Analytics, 2017, 3(2): 121-129.

Sesia M, Katsevich E, Bates S, et al. Multi-resolution localization of causal variants across the genome[J]. Nature Communications, 2020, 11(1):1-10.

Shen X, Ma S, Vemuri P, et al. Challenges and opportunities with causal Discovery Algorithms: Application to Alzheimer's pathophysiology[J]. Scientific Reports, 2020. 10(1):1-12.

Sun J, Bollt E M. Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings[J]. Physica D: Nonlinear Phenomena, 2014, 267: 49-57.

Sun J, Taylor D, Bollt E M. Causal network inference by optimal causation entrop[J]. SIAM Journal on Applied Dynamical Systems, 2015, 14(1):73-106.

Shanghai municipal health commission. Shanghai Special Emergency Response Plan for Public Health Emergencies[Z]. 2020-05-11.

Trave-Massuyes L, Mime R. Gas-turbine condition monitoring using qualitative model-based diagnosis[J]. IEEE Expert, 1997, 12(3): 22-31.

Tu R, Zhang C, Ackermann P, et al. Causal Discovery in the Presence of Missing Data[C]//The 22nd International Conference on Artificial Intelligence and Statistics, 2019: 1762-1770.

Ver Steeg G, Galstyan A. Information transfer in social media[C]//Proceedings of the 21st international conference on World Wide Web. ACM, 2012: 509-518.

Zhou T, Wang Y. Causal relationship inference for a large-scale cellular network[J]. Bioinformatics, 2010, 26(16): 2020-2028.

Zheng X, Aragam B, Ravikumar P, et al. DAGs with NO TEARS: continuous optimization for structure learning[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018: 9492-9503.