Analysis of Conditional Independence Relationship and Applications Based on Layer Sorting in Bayesian Networks

Guofu Xin, Youlong Yang, and Xia Liu

Department of Mathematics, Xidian University, Xi'an, 710071, China
ylyang@mail.xidian.edu.cn

Abstract. Bayesian networks are a probabilistic representation for uncertain relationships, which has proven to be useful for modeling real world problems. Causal Independence and stochastic Independence are two important notations to characterize the flow of information on Bayesian network. They correspond to unidirectional separation and directional separation in Bayesian network structure respectively. In this paper, we focus on the relationship between directional separation and unidirectional separation. By using the layer sorting structure of Bayesian networks, the condition demanded to be satisfied to ensure d-separation and ud-separation hold is given. At the same time, we show that it is easy to find d-separation and ud-separation sets to identify direct causal effect quickly.

Keywords: Directional separation, unidirectional separation, Bayesian network layer sorting.

1 Introduction

Bayesian networks also called Belief Networks or Causal Networks, are a powerful tool for modeling decision-making under uncertainty (see [1-5]). They have been successfully applied to different fields, such as, Machine learning, Prediction and Bioinformatics (see [1, 7-9]). Bayesian networks excel in knowledge representation and reasoning under uncertainty notion of causality suggests an unidirectional separation as graphical representation of causal conditional independence structures. There has been lots of research on theory and application about causal effects in Bayesian networks, for example [4, 6, 10-15]. Pearl's [3] notion of causality suggests a unidirectional separation as graphical representation of causal conditional independence structures. Causal independence allows for defining a measure to characterize the strength of a causal effect of causal networks which is called information flow by Nihay Ay and Daniel Ploani [4]. The cause contribution of A to B imposing S is zero, then we say there is no information flow between A and B after intervening S. The notion of causal effects is based on the possibility to intervene in causal models. An intervention is an action taken to force a variable into a certain state, without reference to its own current state, or the states of any of the other variables. Direct causal effect $A \to B$ is the post-interventional probability

H. Deng et al. (Eds.): AICI 2011, Part III, LNAI 7004, pp. 481–488, 2011.

distribution of B which is defined via mechanisms rather than observations. The interventional formalism (see [3]) provided an appropriate framework for the cause mechanisms in the given system. Causal effects can be identified or post-interventional probability distributions can be calculated by interventional formalism if all variables can be observed. There are three popular figure criteria, front-door, back-door and instrumental variables criteria for causal effects. Pearl's notion [3] of identify causal effect demands to meet front-door criteria or back-door criteria if there are some unobserved variables. Zhao and Zheng [10] have discussed and compared identifiability criteria for causal effects in Gaussian causal models. They have obtained that complete data method is better than front-door, back-door criteria and back-door criteria is better than instrumental variables criteria. These results can offer guidance for choosing better identifiability criterion in practice. Hei and Manabu [11] present an extended set of graphical criteria for the identification of direct causal effects in linear Structural Equation Models (SEMs), they introduce a new set of graphical criteria which uses descendants of either the cause variable or the effect variable as "path-specific" instrumental variables" for the identification of the direct causal effect as long as certain conditions are satisfied.

Directional separation (d-separation for short) is a relation among three disjoint sets of nodes in a directed acyclic graph which is defined in [12]. Two subsets of nodes, X and Y, are said to be d-separated by Z if all chains between the nodes in X and the nodes in Y are blocked by Z. While this condition is characterized by its consistency with stochastic independence structures. Pearl's well known d-separation criterion in a directed acyclic graph is a path separation criterion that can be used to efficiently identify all valid conditional independence relationships in the Markov model determined by the graph. A joint distribution represented by a Bayesian network must satisfy the Markov conditions of the structure: each variable must be independent of its non-descendents given its parents. Unidirectional separation (ud-separation for short) which is defined in [4] is used to judge causal conditional independence. Let A, B, S be three disjoint subsets of nodes. We say that B is ud-separated from A by S if all directed paths from A to B go through S. There exists a directed path from A to B that does not meet S if the information flow between A and B is a positive number after intervene S. A necessary and sufficient condition of ud-separation is given in [4] .What is the relationship between d-separation and ud-separation? How to find the d-separation and ud-separation sets quickly?

The purpose of this paper focus on the relationship between d-separation and ud-separation in directed acyclic graphs. By using layer sorting structure of a Bayesian network, d-separation and ud-separation sets are found quickly to identify direct causal effects .

This paper is organized as follows. In section 2, we introduce background knowledge. In section 3, a new definition which is called layer sorting is introduced, and we discuss a special kind of Bayesian networks, and in this case, both d-separation and ud-separation hold when it satisfy certain condition. In section 4, we introduce two applications about layer sorting.

2 The Basic Concepts and Relation Works

In this section, we introduce some notions and discuss previous relevant works on which our results is based. At the same time, we introduce present relevant background knowledge. We assume that the readers have some basic familiarity with graph theory and Bayesian networks.

A Bayesian network is consist of (1) structure: a directed acyclic graph or a DAG for short, $G=(V,E)$; (2) parameters: represent a joint distribution P over variables. We use higher-case Roman letters for sets of nodes, and lower-case Roman letters for singleton node.

We consider a finite set $V \neq \phi$ of nodes and a set $E \in V \times V$ among these nodes. Such a directed graph $G=(V,E)$, if $\langle v_i, v_j \rangle \in E$, it means that it is a directed edge from v_i to v_j, we note $v_i \to v_j$. If two nodes v_i and v_j, either $v_i \to v_j$ or $v_j \to v_i$, we call v_i and v_j are adjacent, we note $v_i - v_j$. An ordered sequence $\pi = (v_1, v_2, \cdots v_n)$ in a DAG $G=(V,E)$ is called a path from v_1 to v_n, if v_i and v_{i+1} are adjacent for all $i = 1, 2, \cdots n-1$. An ordered sequence $\pi = (v_1, v_2, \cdots v_n)$ in a DAG $G=(V,E)$ is called a directed path from v_1 to v_n, if $v_i \to v_{i+1}$ for all $i = 1, 2, \cdots n-1$. If $v_1 = v_n$, directed path $\pi = (v_1, v_2, \cdots v_n)$ is called a directed cycle. A directed graph with no directed cycles is called a directed acyclic graph.

Let $G=(V,E)$ be a DAG, and x, y, z be three different nodes in G. If $x \to z \to y \in G$, we call z is a serial connection node. If $x \leftarrow z \to y \in G$, we call z is a diverging connection node. If $x \to z \leftarrow y \in G$, we call z is a collider node. A path $\pi = (v_1, v_2, \cdots v_n)$ is called a compound active path given conditioning set Z in DAG G, if each node v_i in the path has one of the two following properties: (1) v_i is not a collider and v_i is not in Z; or (2) v_i is a collider and either v_i or a descendant of v_i in G is in Z. D-separation has been identified as the graphical separation property that is consistent with stochastic conditional independence. It is defined as follows. A path $\pi = (v_1, v_2, \cdots v_n)$ is blocked by a set S, if there is a node $v_j (1 < j < n)$ of the path such that: either $v_j \in S$ and v_j is not a collider, or v_j and all its descendants are not in S, and v_j is a collider. Sets A and B are d-separated given a set S in $G=(V,E)$, if every simple active path between a node in A and a node in B is blocked given conditioning set S. Let $(B \perp_d A \mid S)_G$ denote that A is d-separated from B given S in G. Sets A and B are ud-separated given a set S in $G=(V,E)$, if every directed path from A to B is blocked given conditioning set S. And let $(B \perp_{ud} A \mid S)_G$ denote that B is unidirectional separated from A given a set S in G.

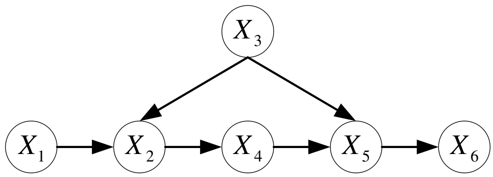

Fig. 1. A directed acyclic graph

For example, we have a set of six nodes $V=\{X_1, X_2, X_3, X_{46}, X_5, X_6\}$, and a set of six edges among these nodes:

$$E=\{\langle X_1, X_2\rangle, \langle X_3, X_2\rangle, \langle X_2, X_4\rangle, \langle X_3, X_5\rangle, \langle X_4, X_5\rangle, \langle X_5, X_6\rangle\},$$

which is shown in Figure 1. The path $X_1 \to X_2 \to X_4 \to X_5 \to X_6$ is direct path. Let $A_1=\{X_6, X_3\}$, $B_1=\{X_1\}$ and $S_1=\{X_4\}$. Then sets A_1 and B_1 are ud-separated by S_1. Let $A_2=\{X_6\}$, $B_2=\{X_1\}$, $S_2=\{X_5\}$. One has that A_2 is d-separated from B_2 given S_2.

3 Main Results and Proof

In this section, we take two parts to discuss the relationship between d-separation and ud-separation.

There is a simple active path between node x and node y given conditioning set Z in G if and only if there is a compound active path between node x and node y given conditioning set Z in G. This implies that simple and compound active paths are interchangeable with respect to the definition of d-separation. Firstly, let us see the following lemma 3.1.

Lemma 3.1. Let $G=(V,E)$ be a DAG, and A, B and S are three disjoint subsets of V, then $(B \perp_d A \mid S)_G$ is the sufficient condition for $(B \perp_{ud} A \mid S)_G$.

Proof. we assume sets A and B are d-separated by S. Let $l=(v_1, v_2, \cdots v_n)$ be an any directed path from the set A to the set B, then l is blocked by S. Each node $v_j (j=1,2\cdots n)$ in l is a serial connection node, according to the definition of d-separation, there must be exist one node $v_j (1<j<n)$ in S. Therefore each directed path between sets A and B goes through S, then B is d-separated from A given S. □

We can see this from Figure 1. Suppose $A=\{X_6, X_3\}$, $B=\{X_1\}$, $S=\{X_4\}$, then B is ud-separated from A given S, but B is not d-separated from A given S.

From lemma 3.1, we know that d-separation is the sufficient condition for ud-separation. In order to discuss the condition that both d-separation and ud-separation hold, we introduce the definition of layer sorting.

Let $G=(V,E)$ be a DAG, F_0 be the set of nodes that has no parents nodes. According to $F_{m+1}=\{v\in V\setminus(F_0,F_1,\cdots F_m): pa(v)\bigcup(F_0\bigcup F_1\bigcup\cdots\bigcup F_m)\neq\phi\}$ where $m=0,1,2\cdots$, we get next layers. Since V is a finite set, for some m, we have $F_{m+1}=\phi$. Therefore, the layers after F_{m+1} are also empty. Assume $K=\max\{m,F_m\neq\phi\}$, we have the disjoint union $V=F_0\bigcup F_1\bigcup\cdots\bigcup F_K$. The corresponding map that $l\to\{0,1\cdots K\}$ assigns to each $v\in V$ its layer number $l(v)$ where $0\le l(v)\le K$.

Suppose $A=\{X_1,X_3\},B=\{X_6\},S=\{X_5\}$, obviously B is d-separated from A given S and it also is ud-separated A given S .

Theorem 3.1. Suppose $G=(V,E)$ be a DAG, and its layer structure $V=F_0\bigcup F_1\bigcup\cdots\bigcup F_K$.

For any non-negative integers a,s,b , then F_b are F_a ud-separated by F_s if $0\le a<s<b\le K$.

Proof. We consider a directed path $\pi=(v_1,v_2,\cdots v_n)$ from F_a to F_b . The corresponding layer numbers are $l(v_1),\cdots l(v_n)$, and $l(v_1)=a$, $l(v_n)=b$. If $l(v_{i+1})>l(v_i)$, we have $l(v_{i+1})=l(v_i)+1$. It implies that there must be one node $l(v_i)=s(a<i<b)$. Therefore, any directed path $\pi=(v_1,v_2,\cdots v_n)$ from F_a to F_b goes through F_s . That is $(F_b\perp_{ud}F_a\mid F_s)_G$ holds. □

Theorem 3.2. Suppose $G=(V,E)$ be a DAG, and its layer structure

$$V=F_0\bigcup F_1\bigcup\cdots\bigcup F_K .$$

If the parents of F_m are included in $F_0\bigcup F_1\bigcup\cdots\bigcup F_m$, for three non-negative integers a,s,b , if $0\le a<s<b\le K$, F_b and F_a are ud-separated and d-separated by F_s .

Proof. Firstly, we prove that $\langle F_a\perp_d F_b\big|F_s\rangle_G$ holds. Let $l=(v_1,v_2,\cdots v_i,\cdots v_j,\cdots v_k,\cdots v_q)$ be a simple active path from F_a to F_b. There exist one nodes of F_s at least is not a collider in path l . Assume $p=\{v_i\in l\}\bigcap F_s$ where the path l goes through F_s. If one node in p is not a collider, then the conclusion that $(F_a\perp_d F_b\mid F_s)_G$ is right. Otherwise, all nodes in p are colliders in the path l . We consider v_k , since $v_k\in p$, then $v_{k+1}\notin F_s$. Then the subpath $l^{'}=(v_{k+1},v_{k+2},\cdots v_q)$ of l is a simple active path from F_{s-1} to F_b. According to lemma 3.2, there exists one node in $l^{'}$ goes through F_s . This is contradictory to the definition on p . Therefore $(F_a\perp_d F_b\mid F_s)_G$ holds.

Because all directed paths from F_a to F_b go through F_s, according to theorem 3.1, $(F_b \perp_{ud} F_a \mid F_s)_G$ holds. □

In that special kind of Bayesian networks, there is one node on each active path from F_a to F_b which is not a collider in conditioning set F_s.

Let $G=(V,E)$ be a DAG. Further let A, B and S are three disjoint subsets of V. If $(A \perp_d B \mid S)_G$ and $(B \perp_{ud} A \mid S)_G$ hold simultaneously, any active path l from A to B must meet one of the following two conditions.

Either l is a directed path in itself, then l must go through S, or l is not a directed path in itself, (1) l does not go through S, there must be a collider node in l and its descendents are not in S, (2) l goes through S, let $(s_1, s_2, \cdots s_k)$ be the nodes that l goes through S, if $k=1$, node s_1 is not a collider node in S. If $k \geq 2$, there is a node $s_i (1 \leq i \leq k)$ which is not a collider in l.

Because an active path may be also a directed path in itself, and may be a directed path is a sub-path of an active path. In order to discuss the condition demanded to be satisfied to ensure d-separation and ud-separation hold simultaneously, we discuss it in two special cases, any other cases seems to be these two cases.

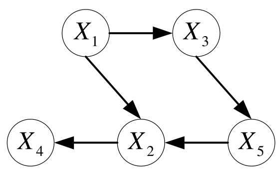

Fig. 2. Two special cases to judge d-separation and ud-separation

Case1: Let $A_1=\{X_1\}, B_1=\{X_5\}, S_1=\{X_3\}$ in Figure 2. Then $(B_1 \perp_d A_1 \mid S_1)_G$ and $(B_1 \perp_{ud} A_1 \mid S_1)_G$ hold simultaneously. Namely, each directed path from A_1 to B_1 goes through S_1. If one has a active paths from A_1 to B_1 which not go through S_1, there is a collider node in these active paths and its descendents are not in S_1.

Case 2: Assume $A_2=\{X_1\}, S_2=\{X_2, X_3\}, B_2=\{X_4, X_5\}$ in Figure 2. Then B_2 is ud-separated from A_2 by S_2 in G. However B_2 is not d-separated from A_2 by S_2 because nodes X_1 and X_5 are not independence given X_2 in path $X_1 \rightarrow X_2 \leftarrow X_5$. In the case, if there is an active path from A_2 to B_2 which has only a collider node in S_2, the case is not right.

4 Applications and Discussion

In this section, we discuss two applications of layer sorting. One is that we can get a Bayesian network's topological sequence using layer sorting. The other is we can get d-separation and ud-separation sets to indentify the direct causal effects.

Assume we give a causal graph G as is shown in Figure 3 and its layer structure is shown in Figure 4.

From Figure 4, one has $F_0 = \{A, B\}, F_1 = \{C, D, E\}, F_2 = \{F, H\}, F_3 = \{G\}$. Firstly, we choose one node from F_0 and delete it until F_0 is empty. Secondly, we select one node that has not parents from $V \setminus F_0$ and delete it. Do like this until all nodes are opted. The topological sequence of causal graph G is

$$(A, B, C, D, E, F, G, H).$$

Let F and H be two different nodes in G. We want to judge whether the causal effect of F on H is identifiable. Using layer sorting, we obtain its layer structure which is shown in Figure 4. From Figure 4, we have F and H are in the same layer. We know the d-separation set is consist of C and D. We use this example to show that we can get the d-separation and ud-separation sets easily using layer sorting for some special kind of Bayesian Networks.

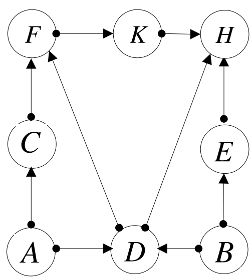

Fig. 3. A causal graph G

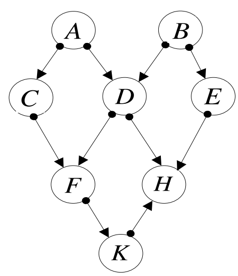

Fig. 4. The layer structure of G

The relationship between d-separated and ud-separated is discussed in this paper, we obtain that d-separated is sufficient condition for ud-separated. By using the layer sorting structure of Bayesian networks, the condition demanded to be satisfied to ensure d-separation and ud-separation hold is given. D-separation and ud-separation play a very important role in the indentify of direct causal effects. In the next time, we will use layer sorting together with d-separation and ud-separation's nature to indentify the direct causal effects.

References

1. Friedman, N.: Inferring Cellular Networks Using Probabilistic Graphical Models. Science 303(5659), 799–805
2. Neapolitan, R.E.: Learning Bayesian Networks. Prentice-Hall, Englewood Cliffs (2003)
3. Pearl, J.: Causality: Models, Reasoning and Inference. Cambridge University Press, Cambridge (2000)
4. Ay, N., Polani, D.: Information Flows in Causal Networks. Advances in Complex Systems 11(1), 17–41 (2008)

5. Chickering, D.M., Meek, C.: On the incompatibility of faithfulness and monotone DAG faithfulness. Artificial Intelligence 170(8), 653–666 (2006)
6. Lauritzen, S.L.: Graphical Models for Causal Inference. Royal Economics Society Summer School, Oxford, Lecture Notes (2000)
7. You-long, Y., Yan, W.: VC dimension and inner product space induced by Bayesian networks. International Journal of Approximate Reasoning 50(7), 1036–1045 (2009)
8. Slezak, D.: Degrees of conditional dependence: a framework for approximate Bayesian networks and examples related to the rough set-based feature selection. Information Sciences 179(3), 197–209 (2009)
9. Yang, Y., Wu, Y.: Inner Product Space and Concept Classes Induced by Bayesian Networks. Acta Application Mathematicae 106(3), 337–348 (2009)
10. Zhao, H., Zheng, Z.-G.: Comparing Identifiability Criteria for Causal Effects in Gaussian Causal Models. Acta Mathematica Scientia 28A(5), 808–817 (2008)
11. Chan, H., Kuroki, M.: Using Descendants as Instrumental Variables for the Identification of Direct Causal Effects in Linear SEMs. In: International Conference on Artificial Intelligence and Statistics (AISTATS), Chia Laguna Resort, Sardinia, Italy (2010)
12. Levitz, M., Perlman, M.D., Madigan, D.: Separation and completeness properties for AMP chain graph Markov Models. The Annals of Statistics 29(6), 1751–1784 (2001)
13. Tian, J., Pearl, J.: On the testable implications of cause models with hidden variables. In: Proceedings of the Eighteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI 2002), pp. 519–527 (2002)
14. Geng, Z., He, Y.-B., Wang, X.-L.: Relationship of causal effects in a causal chain and related inference. Science in China 47A, 730–740 (2004)
15. Tian, J., Pearl, J.: On the identification of cause effects, Technical report 475290-L, Tech. Rep. Cognitive Systems Laboratory, University of California at Los Angeles (2003)