에센셜 그래프를 바탕으로 한 격자 조건부 독립 모델†

(Lattice Conditional Independence Models Based on the Essential Graph)

김 주 성*, 서 민 자**, 윤 명 영***
(Ju Sung Kim*, Min Ja Seo**, Myoong Young Yoon***)

요 약 결측치가 존재하는 비 단조형 데이터에 대한 패턴 분석과 비 내포형 종속 회귀 모형 분석에 격자 조건부 독립 모델이 최근 도입되고 있다. 이러한 접근 방법은 데이터 패턴 분석에 성공적으로 적용되고 있지만 격자 조건부 독립 모델을 찾는 계산적 부담이 따른다. 본 논문에서는 이러한 단점을 극복하기 위하여 에센셜 그래프를 바탕으로 격자 조건부 독립 모델(LCIM)을 찾는 새로운 방법을 제안한다. 또한, LCIM 클래스가 특정한 비 순환 방향 그래프 모델과 마르코프 동등한 모든 추이적 비 순환 방향 그래프의 모델 클래스와 일치함을 밝혔다.

핵심주제어 : 비순환 방향 그래프, 격자 조건부 독립, 마코프 성질, 에센셜 그래프

Abstract Recently, lattice conditional independence models(LCIMs) have been introduced for the analysis of non-monotone missing data patterns and of non-nested dependent regression models. This approach has been successfully applied to solve various problems in data pattern analysis, however, it suffers from computational burden to search LCIMs. In order to cope with this drawback, we propose a new scheme for finding LCIMs based on the essential graph. Also, we show that the class of LCIMs coincides with the class of all transitive acyclic directed graph(TADG) models which are Markov equivalent to a specific acyclic directed graph(ADG) models.

Key Words : Acyclic directed graph(ADG), Markov property, Essential graph

1. Introduction

Many important statistical models involve Markov type processes, meaning that some conditional statistic for a data, given values at other samples, depends only on those near the point of interest. Lattice conditional independence models (LCIMs) were introduced by Andersson and Pearman for the analysis of non-monotone multivariate missing data patterns and non-nested dependent linear regression models[1]. Especially, graphical Markov models determined by conditional independence are their

† 이 논문은 충북대학교 2004학년도 학술연구지원 사업의 연구비 지원에 의하여 연구되었음.
* 충북대학교 정보통계학과 교수
** 충북대학교 정보통계학과 강사
*** 충청대학 컴퓨터학부 부교수

appropriateness for characterizing the state in a wide range or physical systems. Undirected graphs have found applications as models of dependence for spatial stochastic processes, image analysis and contingency tables.

In LCIM, conditional independence relations among the variables are specified by the intersection properties of a finite distributive lattice which. The finite distributive lattice is presented as a ring of subsets of a finite index set of variables[2].

Graphical Markov models determined by acyclic directed graphs(ADGs) admit especially simple statistical analysis and there may be several ADGs that determine the same Markov model. The family of all ADGs with a given set of vertices is naturally partitioned into Markov equivalence classes, each class being associated with a unique statistical model[3]. Since the order $|J(L)|$ of the set $J(L)$ which consist of all non-zero join-irreducible element of lattice L smaller than that of the lattice L, the TADG representation is more economical than the LCI representation. So, we show that the class of LCIM models coincides with the class of TADG models. If a specific ADG G is transitive, G determines a statistical model that Markov equivalent to some LCIMs. Although specific ADG G may not be transitive, its Markov equivalence class may contain a transitive ADG in which case G is Markov equivalent to some LCIM[4, 5]. Thus, in order to determine whether a specific ADG G is Markov equivalent to some LCIM, we must find whether Markov equivalence class $[G]$ contain at least one transitive ADG [6, 7].

Since $[G]$ can be exponentially large, exhaustive search of $[G]$ is computationally unfeasible for large graphs. In order to cope with this drawback, in this paper, we provide a method, based on the essential graph G^* of G, of finding TADG which is Markov equivalent to a specific ADG G. From this method, we can easily determine whether a specific ADG G is Markov equivalent to some lattice conditional independence model.

2. LCIMs and Graphical Markov models

Our terminology and notation follows Graphical Models in Applied Multivariate statistics by Whittaker [8] and General Lattice Theory by George Grätzer[9].

Let $L \equiv L(\wedge, \vee)$ be a finite distributive lattice with minimum 0. The set $J(L)$ of non-zero join- irreducible elements of L is defined as:

$$J(L) = \{x \in L \mid x \neq 0,\ x = y \vee z \Rightarrow x = y \text{ or } x = z\}.$$

Then $(J(L), \leq)$ is a finite partially ordered set(poset) under partial ordering $\leq$ inherited from L : $x \leq y$ iff $x \wedge y = x$. A subset A of $J(L)$ is ancestral if $x \in A$ whenever $x \in J(L)$, $x \in J(L)$, $y \in A$ and $x \leq y$. From now, we consider multivariate probability distribution functions f on a product probability space $X = (X_i \mid i \in K)$, where K is a finite index set and X_i are random variables.

Definition 2.1 Let L be a finite distributive lattice.

(1) A probability function f on X is said to satisfy the lattice conditional independence property(LCIP) relative to L if, for every $x,\ y \in L$,

$$an(x) \perp an(y) \mid an(x \wedge y) \quad \text{, where}$$
$$an(x) = \{z \in L \mid z \leq x\}.$$

(2) The set $L_X(L)$ of all probability function f on X is said to satisfy the LCIP relative to L is called the lattice conditional independence model determined by L. Birkhoff's theorem states that the mapping:

$$L \rightarrow A((J(L), \leq))$$
$$x \rightarrow an(x) = \{y \in J(L) \mid y \leq x\}$$

determines a lattice isomorphism between finite distributive lattice L and the ring $A((J(L), \leq))$.

Definition 2.2 A graph G is a pair (K, E), where K is a finite set of vertices and a set of edges, E, consisting of pairs of elements taken from K, denoted by $G=(K,E)$. If G is contains only directed edges, it is called a directed graph(digraph). A graph G is called a chain graph if it does not contain any directed cycles. A chain graph that is also a digraph is called an acyclic digraph(ADG).

In our discussion for the following definitions, theorems and lemmas, we consider a vector of variables $X=(X_1,\dots,X_k)$ and the corresponding set of vertices $K=\{1, 2, \dots, k\}$.

Definition 2.3 Let $G=(K,E)$ be an ADG. A probability function f on X is said to satisfy:
(1) the local Markov property(LMP) relative to G if, for every $k\in K$, $X_k \perp X_b \mid X_a$, where $a=bd(k)$, $b=K-(\{k\}\cup a)$.
(2) the global Markov property(GMP) relative to G if, for all disjoint subsets a, b and c of k $X_b \perp X_c \mid X_a$ whenever b and c are separated by a.
(3) the well-numbered Markov property(WNMP) relative to G if, for each $i=2, \dots, k$, $k_i \perp \{k_1,\dots,k_{i-1}\}-pa(k_i)$ where $k=|K|$ and $k_1,\dots,k_k$ is well-numbering of members of K.

Definition 2.4 Let $G=(K,E)$ be an ADG.
(1) The set of $M_X(G)$ of all probability functions f on X that satisfy the three equivalent Markov properties GMP, LMP and WNMP relative to G is called the Markov model determined by the ADG G or simply the ADG model determined by G.
(2) An subset $R(\subset K)$ is called ancestral in G if $r_1\in R$ whenever $r_1\in K$, $r_2\in R$, and $r_1\le r_2$ in G. Ancestral ring $R(G)$ is defined to collection of all ancestral subsets of G under union and intersection.
(3) A probability function f on X is said to satisfy the lattice conditional independence property(LCIP) relative to G if, for every pair $X_1, X_2 \in R(G)$, $X_1 \perp X_2 \mid X_1\cap X_2$ where $R(G)$ is called ancestral ring. (4) The set $L_X(G)$ of all probability functions f on X that satisfy the LCIP relative to G is called the lattice conditional independence model determined by the ADG G.
Let $G=(K,E)$ be an ADG. For any probability density function f on X, then $M_X(G)\subseteq L_X(G)$.

Definition 2.5 Let $G=(K,E)$ be an ADG. G is said to be a transitive ADG(TADG) if $\alpha\to\beta\in G$ and $\beta\to\gamma\in G$ then $\alpha\to\gamma\in G$ where α, β, $\gamma\in G$ and $\to$ stands for directed edge. Note that the ring $A((J(L),\le))$ of all ancestral subsets of the poset $(J(L), \le)$ is identical to the ancestral ring $A((J(L), E^{<}))$, where $(J(L), E^{<})$ is the TADG given by
$E^{<} = \{(x, y)\in J(L)\times J(L) \mid x<y\}$.

Lemma 2.6 Let $G=(K,E)$ be an TADG. For any probability density function f on X, then GMP $\Leftrightarrow$ LMP $\Leftrightarrow$ WNMP and $M_X(G) = L_X(G)$.

Theorem 2.7 The class of LCI models coincides with the class of TADG models.
Proof. In the first, we want to show that a LCI model can be expressed as the TADG model. Let $L_X(L)$ be a LCI model, where $X=(X_1,\dots,X_n)$ for $1,\dots,n\in J(L)$. Then by Definition 2.1 2.4 and lattice isomorphism, $L_X(L)=L_X((J(L), E^{<}))$ and $L_X((J(L), E^{<}))=M_X((J(L), E^{<}))$ by Lemma 2.6. Thus, the LCI model $L_X(L)$ can be expressed as the TADG model $M_X((J(L), E^{<}))$.

In the second, we want to show that a TADG model can be expressed as the LCI model. Consider a TADG model $M_X(G)$, where $G=(K,E)$ is a TADG and $X=(X_1,\ldots,X_n)$ for $1,\ldots,n\in K$. Then by Lemma 2.6, $L_X(G)=M_X(G)$. And by Definitions 2.1 and 24 $M_X(G)=L_X(R(G))$. Thus, $M_X(G)$ can be expressed as the LCI model $L_X(R(G))$.

Remark 2.8 In general $|\mathcal{J}(L)|$ is smaller than $|\mathcal{J}|$, so, the TADG representation is more economical than the LCI representation.

3. Markov equivalence and essential graph

In this section, we define Markov equivalent for two ADGs G_1 and G_2. Two ADGs G_1 and G_2 are Markove equivalent on X indexed by K if $M_X(G_1)=M_X(G_2)$. The Markov equivalence class $[G]$ is defined as follows: $[G]=\{G_i \mid G_i\cong G\}$ [7]. Two ADGs are Markov equivalent if and only if they have the same skeleton and the same immoralities. The essential graph G^* associated with G is graph $G^*=\cup\{G_i \mid G_i\cong G\}$. The following theorem is sufficient and necessary conditions for the essential graph G^* of graph G.

Theorem 3.1 A graph $G=(K,E)$ is equal to G^* for some ADG G if and onlyif G satisfies the following four conditions:

(1) G is a chain graph;

(2) For every chain component ξ of G is chordal;

(3) The configuration $x\rightarrow y-z$ does not occur as an induced sub-graph of G ;

(4) Every arrow $x\rightarrow y\in G$ is strongly protected in G. A digraph G' is acyclic and equivalent to the ADG G if and only if G' is obtained from G^* by orienting the edges of each chain component G^*_ξ in any acyclic and moral way. The G is Markov equivalent to its essential graph G^*. By using the following theorem we can find essential graph G^* associated with an ADG G.

Theorem 3.2 Let $G=(K,E)$ be a ADG. Define $G_0=G$. For $i\geq 1$, convert every arrow $x\rightarrow y\in G_{i-1}$ that is not strongly protected in G_{i-1} into a line $x-y$, obtaining a graph G_i. Stop after l steps, where $l\geq 0$ is smallest non-negative integer such that $G_l=G_{l+1}$. Then $l\leq |K|$ and $G_l=G^*$. This requires at most $O(|K|n^4)$ operations. The following theorem leads to a feasible computation for deciding whether an ADG G is Markov equivalent to some LCI model.

Theorem 3.3 Let $G=(K,E)$ be an ADG is Markov equivalent to some LCIM if and only if none of the following five configuration occur as an induced sub-graph of the associated essential graph G^*:

(1) $x\rightarrow y\rightarrow z$

(2)
$$\begin{matrix} x & \rightarrow & y \\ | & & \uparrow \\ z & - & w \end{matrix}$$

(3) $x\leftarrow y-z\rightarrow w$

(4) $x-y-z\rightarrow w$

(5) $x-y-z-w$

To the proof of Theorem, we introduce a Lemma for orienting the undirected edges to directed edges of a essential graph $G^*=(K,E^*)$.

Lemma3.4 For each vertex $k\in K$, define the degree $d(g)$ of $g\in G^*$ as $d(g)=|\{g'\in K \mid g-g'\in G^* \text{ or } g\rightarrow g'\in G^*\}|$. For each chain component $\xi\in\Xi(G^*)$ such that $|\xi|\geq 2$, if orient the (necessarily undirected) edges of

G^*_ξ as follows:

Step 1) Define

$$x_1 = x_1(\xi) = \arg\max\{d(x_1) \mid x_1 \in \xi\}$$

Step 2) Define

$$\xi_1 = \xi_1(\xi) = \{x' \in \xi \mid x_1 - x' \in G^*_\xi\} \cup \{x_1\}$$

Step 3) For every $x' \in \xi_1 - \{x_1\}$, orient the edge $x_1 - x'$ as $x_1 \to x'$.

For $i \geq 2$,

Step 1) Define

$$x_i = x_i(\xi) = \arg\max\{d(x_i) \mid x_i \in (\xi_1 \cup \ldots \cup \xi_{i-1}) - \{x_1, \ldots, x_{i-1}\}\}$$

Step 2) Define

$$\xi_i = \xi_i(\xi) = \{x' \in \xi \mid x_i - x' \in G^*_\xi\} - \{x_1, \ldots, x_{i-1}\}$$

Step 3) For every $x' \in \xi_i$, orient the edge $x_i - x'$ as $x_i \to x'$.

Then, after k-step,

(1) $x_1, \ldots, x_i$ are distinct vertices in ξ.

(2) $x_1 \in \xi_1$ and $x_i \in \xi_1 \cup \ldots \cup \xi_{i-1} \subseteq \xi$ for $2 \leq i \leq k$

(3) $k \leq |\xi|$. where $k = \min\{i \mid \{x_1, \ldots, x_i\} = \xi_1 \cup \ldots \cup \xi_i\}$.

From the above Lemma, we can see that the digraph G'_ξ is acyclic, the ADG G'_ξ is moral by the orientation algorithm.

Proposition 3.2 (1) $d(x_1) \geq d(x_i)$ for $i \geq 2$. (2) If the configuration $x_i \to x_m \to x_j$ and $x_i \to x_j$ occurs in G'_ξ, then $d(x_m) \geq d(x_j)$.

Proof: (1) is obvious. For (2), by definition of x_m it suffices to show that $x_j \in \{(\xi_1 \cup \ldots \cup \xi_{m-1}) - \{x_1, \ldots, x_{m-1}\}\}$. This is a consequence of the following three observations:

(a) $x_i \to x_j \in G'_\xi \Rightarrow x_j \in \xi_i$;

(b) $x_i \to x_m \in G'_\xi \Rightarrow i < m \Rightarrow \xi_i \subseteq \xi_1 \cup \ldots \xi_{m-1}$;

(c) $x_m \to x_j \in G'_\xi \Rightarrow m < j \Rightarrow x_j \notin \{x_1, \ldots, x_{m-1}\}$

Let $G = (K, E)$ be an ADG and G' be the digraph obtained by applying the Lemma3.4 to the essential graph G^* of G. Then G' is acyclic and equivalent to G.

Proof of theorem 3.3: ($\Leftarrow$) If any of the five configurations (1)-(5) occur as an induced subgraph of G^*, then every orientation of the undirected edges of G^* must produce a non-transitive digraph, hence every ADG in $[G]$ is non-transitive.

($\Rightarrow$) If none of these configurations occur in G^*, suppose that G' is non-transitive. Then at least one of the following four non-transitive configurations must occur as an induced subgraph of G':

(a) $x \; \alpha \; y \; \alpha \; z$ (b) $x \; \alpha \; y < z$

(c) $x < y \; \alpha \; z$ (d) $x < y < z$

where α stands for essential arrow, $<$ stands for non-essential arrow. Since (a) would be violate the nonoccurrence of (1) in G^*, clearly (a) cannot occur in G'. Since G^* would fail to ratify condition (3) of Theorem 3.1, (b) cannot occur in G'. Suppose that (c) occurs in G'. We assert that $A - B = \emptyset$, where

$$A = \{x' \in K \mid x - x' \in G^* \; or \; x \; \alpha \; x' \in G^*\}$$
$$B = \{x' \in K \mid y - x' \in G^* \; or \; y \; \alpha \; x' \in G^*\}$$

If $A - B = \emptyset$, then each $x' \in A - B$ must occur in one of the following two configurations as an induced sub-graph of G^*:

$$\text{(e)} \quad \begin{matrix} y & \alpha & z \\ \wedge & & \text{x} \\ x & < & x' \end{matrix} \qquad\qquad \text{(f)} \quad \begin{matrix} y & \alpha & z \\ \wedge & & \text{x} \\ x & \alpha & x' \end{matrix}$$

where x indicates either no edge, an undirected edge, or a directed edge of unspecified orientation. The absence of any edge between x' and z in the above (e) or (f) would contradict the non-occurrence in G^* of (3) and (4), respectively. The occurrence of x' a z in (e) or (f) would contradict the non-occurrence in G^* of (1) and (2), respectively. Finally the non-occurrence of z a x' in (1) or (2) would contradict non-occurrence in G^* of (1). Thus $A-B=\emptyset$. Because $z \in B-A$, it follows that $d(y) \equiv |B| > |A| \equiv d(x)$. Since $x-y \in G^*$, $\{x,y\} \subseteq \xi$ for some $\xi \in \Xi(G^*)$. If $x=x_1=x_1(\xi)$, then $d(x) \geq d(y)$. If $x=x_i=x_i(\xi)$ for some $i \geq 2$, then there exist $j > i$ such that $x_j > x \in G'_{\xi}$. Therefore $x_j \in A \subseteq B$, hence either $x_j > y \in G'_{\xi}$, $y > y_j \in G'_{\xi}$ or $y\ a\ x_j \in G'_{\xi}$. Since $x > y \in G'_{\xi}$, the second and third possibilities would violate the acyclicity of G', hence the first possibility must hold. This implies that $d(x) \geq d(y)$, again a contradiction. Thus (c) cannot occur in G'. Lastly, suppose that (d) occur in G'. we again assert that $A-B=\emptyset$. If $A-B \neq \emptyset$, then each $x' \in A-B$ must occur in one of the following two configurations as an induced subgraph of G^*

	y	$<$	z		y	$<$	z
(g)	$\wedge$		x	(h)	$\wedge$		x
	x	$<$	x'		x	a	x'

The absence of any edge between x' and z in these two configurations would contradict the non- occurrence in G^* of (4) and (5), respectively.

If x' a z in (g) and (h), then this would contradict conditions (2) and (3) ,respectively, of Theorem 3.1. The occurrence of $x' < z$ in (g) and (h) would contradict both the acyclicity of G^* and the condition (3) of Theorem 3.1. Finally, the occurrence of $z < x'$ in (g) would contradict both acyclity and condition (3), while its occurrence in (h) would contradict the non- occurrence of (2) in G^*. Thus, $A-B=\emptyset$.

Because $z \in A-B$, $d(y) > d(x)$. Here $\{x,y,z\} \subseteq \xi$ for some $\xi \in \Xi(G^*)$. However we can show that $d(x) \geq d(y)$, exactly as above, a contradiction. Thus (d) cannot occur in G'. This complete the proof of Theorem.

Examples

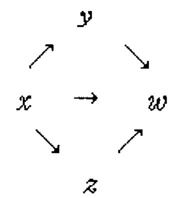


Figure 1. A TADG $G=(K,E)$ with vertex set $K=\{x,y,z,w\}$. The model $M_X(G)$ specifies the single conditional independence $y \perp z \mid x$. Ancestral ring $R(G)=\{\emptyset, \{x\}, \{x,y\}, \{x,y,z\}, \{x,y,z,w\}\}$

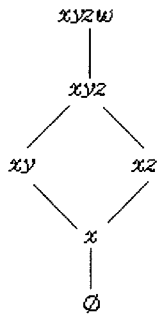


Figure 2. A finite distributive lattice L with vertex set $\{x,y,z,w\}$.

The LCI model $L_X(L)$ specifies the conditional independence $xy \perp xz \mid x$. The join-irreducible set $J(L)=\{x, xy, xyz, xyzw\}$. $(J(L), E^{<})$is isomorphic to TADG G and L is identical to $R(G)$ in Figure 1.

4. Concluding remarks

Many important statistical models involve Markov type processes, meanning that some conditional statistics for a data, given values at other samples, depends only on those near the point of interest. Especially, lattice conditional independence models have tractable properties and precisely suited for the analysis of non-monotone missing data patterns and a family of non-nested dependent univariate regression models. In this paper, we proposed a method that based on the essential graph G^* of those G such that $[G]$ contains at least one transitive ADG and therefore is Markov equivalent to some LCIM. Our future works will be concerned with the lattice conditional independent model determined by countably distributive lattice.

References

[1] S.A. Andersson, M.D. Perlman. "Lattice models for conditional independence in a multivariate normal distribution" Anals statiatic, vol21(3) pp 1318-1358,1993.

[2] S.A. Andersson, D. Madigan, and M.D. Perl man, "A characterization of Markov equivalence classes for acyclic digraph," Anals statistic, vol.3(4), pp.567-575, 1996.

[3] S.A. Andersson and M.D. Peralman, "Alternative Markov properties for chain graphs," Scandinavian Journal of Statistics, vol.28, pp. 33-85, 2001.

[4] S.A. Andersson, "The lattice structure of orthogonal linear models and othogonal variance component models," Scand. J. Statistic, vol.17, pp.287-319, 1990.

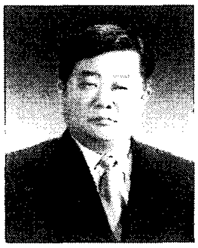
[5] R. Auvray and L.Wehenkel, "On the construction of the inclusion boundary neighborhood for Markov equivalence classes Bayesian network structures," In ptoc. of the 18 Conf. on Uncertainty in art. vol.1, pp.987-1001, 2002.

[6] P. Dawid and S. Lauritzen, "Hyper-Markov laws in the statistical anslysis of decomposable graphical models," Annals of Statistics, vol.21, pp.1272-1317, 1993.

[7] G. Gierz, K. Hofmann, K. Keimel, and M. Mislove, "A Compendium of Continuous Lattices." Springer-Verlag Berlin Heidelberg.

[8] J. L. Whittacker, "Graphical models in applied multivariate statistics," Willy, NY, 1990.

[9] George Grätzer. "General Lattice Theory", Prentice-Hall, Inc., 1988.

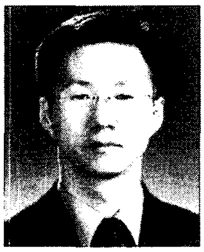


김 주 성 (Ju Sung Kim)

1976년 충남대학교 수학과 졸업(학사)
1978년 충남대학교 대학원 수학과(이학석사)
1991년 경북대학교 대학원 통계학과(이학박사)
1990년~현재 한국통계학회 평의원
1993년~현재 통계청 자문위원
1981년~현재 충북대학교 정보통계학과 교수
(관심분야: 퍼지이론, 패턴분류, 다변량, 회귀분석)

서 민 자 (Min Ja Seo)

1985년 충북대학교 수학과 졸업(학사)
1987년 충북대학교 대학원 수학과(이학석사)
2000년 충북대학교 대학원 통계학과(박사수료)
1988년~현재 충북대학교, 서원대학교 강사
(관심분야: 퍼지이론, 다변량, 회귀분석, 토폴로지)



윤 명 영 (Myoong Young Yoon)

1984년 충북대학교 계산통계학과 (이학사)
1987년 충북대학교 대학원 계산통계학과(이학석사)
1997년 충북대학교 대학원 전자계산학과(이학박사)
1987년~1990년 육군 전산장교 근무
1997년~1999년 충청대학 정보센터 원장
1998년~현재 충청북도 기술자문위원
1999년~현재 충청대학 중소기업종합지원센터 소장
1992년~현재 충청대학 컴퓨터학부 부교수
(관심분야: 화상처리, 패턴인식, 데이터 마이닝)