Axiomatizing Conditional Independence and Inclusion Dependencies

Miika Hannula

University of Helsinki

6.3.2014

Introduction

- We provide an axiomatization for the so-called conditional independence atoms (which correspond to EMVDs + FDs in database theory) and inclusion atoms taken together (a joint work with Kontinen).
- This work arises from the dependence logic/lax team semantics framework.

Axiomatizating Dependence Logic

Dependence Logic ($\mathcal{D}$) cannot be axiomatized: validity problem of $\mathcal{D}$-formulas is of the same complexity as that of full second-order logic formulas.

Axiomatizating Dependence Logic

Dependence Logic ($\mathcal{D}$) cannot be axiomatized: validity problem of $\mathcal{D}$-formulas is of the same complexity as that of full second-order logic formulas.

- We can axiomatize fragments of $\mathcal{D}$ (or variants) for which the logical consequence is (at least) semidecidable: for instance, first-order consequences of $\mathcal{D}$-sentences (Kontinen and Väänänen, 2012) and of $\mathrm{FO}(\perp_{\mathrm{c}})$-sentences (H. 2013).

Axiomatizating Dependence Logic

Dependence Logic ($\mathcal{D}$) cannot be axiomatized: validity problem of $\mathcal{D}$-formulas is of the same complexity as that of full second-order logic formulas.

- We can axiomatize fragments of $\mathcal{D}$ (or variants) for which the logical consequence is (at least) semidecidable: for instance, first-order consequences of $\mathcal{D}$-sentences (Kontinen and Väänänen, 2012) and of $\mathrm{FO}(\perp_c)$-sentences (H. 2013).

Axiomatizations of dependence atoms (+variants):

- Dependence atoms (Armstrong axioms for FDs, 1974),
- Pure independence atoms (Geiger, Paz and Pearl for probabilistic independence, 1991),
- Inclusion atoms (Casanova, Fagin and Papadimitrou for INDs, 1982).

Conditional Independence Atom

A *conditional independence atom* $\vec{y} \perp_{\vec{x}} \vec{z}$ is defined for sequences of variables $\vec{x}$, $\vec{y}$, and $\vec{z}$ (which are not necessarily disjoint or of the same length). Its team semantics is defined as follows:

TS-ind: $X \models \vec{y} \perp_{\vec{x}} \vec{z}$ if and only if for any two $s, s' \in X$ which assign the same value to $\vec{x}$ there exists a $s'' \in X$ which agrees with s with respect to $\vec{x}$ and $\vec{y}$ and with s' with respect to $\vec{z}$.

Conditional Independence Atom cont.

Let X be a team and $s, s' \in X$ such that $s(x) = s'(x')$, and assume that $\mathcal{M} \models_X y \perp_x z$.

	x	y	z	v
s	5	2	4	0
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
s'	5	1	3	1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Conditional Independence Atom cont.

Let X be a team and $s, s' \in X$ such that $s(x) = s'(x')$, and assume that $\mathcal{M} \models_X y \perp_x z$. Then we find $s'' \in X$ as in the picture:

	x	y	z	v
s	5	2	4	0
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
s'	5	1	3	1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
s''	5	2	3	2

Embedded Multivalued Dependency

Conditional independence atom generalizes embedded multivalued dependency (EMVD) of database theory. For sets of attributes X, Y, Z, $X \twoheadrightarrow Y|Z$ is an EMVD, given the following semantic rule

- $r \models X \twoheadrightarrow Y|Z$ if and only if for every $t_0, t_1 \in r$ with $t_0[X] = t_1[X]$ there is $t_2 \in r$ such that $t_2[X] = t_0[X]$, $t_2[Y] = t_0[Y]$ and $t_2[Z \setminus XY] = t_1[Z \setminus XY]$.

Hence $X \twoheadrightarrow Y|Z$ corresponds to the independence atom $Y \perp_X Z \setminus XY$.

Functional Dependency

A functional dependency (FD) $X \to Y$ which corresponds to the dependence atom $=\!(X, Y)$ can also be expressed as the conditional independence atom $Y \perp_X Y$.

	X	Z	Y	V
s	5	2	2	4
s'	4	6	3	7
s''	5	9	2	4
s'''	4	1	3	5

Functional Dependency

A functional dependency (FD) $X \to Y$ which corresponds to the dependence atom $=\!(X, Y)$ can also be expressed as the conditional independence atom $Y \perp_X Y$.

	X	Z	Y	V
s	5	2	2	4
s'	4	6	3	7
s''	5	9	2	4
s'''	4	1	3	5

Hence, conditional independence atoms = EMVDs + FDs

Inclusion Atom

An *inclusion atom* $\vec{x} \subseteq \vec{y}$ is defined for sequences of variables $\vec{x}$ and $\vec{y}$ of the same length (both possibly with repetitions). Its semantic rule is given as follows:

TS-inc: $X \models \vec{x} \subseteq \vec{y}$ if and only if for all $s \in X$ there is $s' \in X$ such that $s(\vec{x}) = s'(\vec{y})$.

Axiomatizing $\{\perp_{\mathrm{c}}, \subseteq\}$-atoms

Given a set $\Sigma \cup \{\phi\}$ of $\{\perp_{\mathrm{c}}, \subseteq\}$-atoms, the implication problem $\Sigma \models \phi$ is to decide whether $X \models \phi$ for all teams X such that $X \models \Sigma$. We will present a complete axiomatization of the implication problem for $\{\perp_{\mathrm{c}}, \subseteq\}$-atoms.

Axiomatizing $\{\perp_c, \subseteq\}$-atoms

Given a set $\Sigma \cup \{\phi\}$ of $\{\perp_c, \subseteq\}$-atoms, the implication problem $\Sigma \models \phi$ is to decide whether $X \models \phi$ for all teams X such that $X \models \Sigma$. We will present a complete axiomatization of the implication problem for $\{\perp_c, \subseteq\}$-atoms.

Since the implication problem for EMVDs is undecidable (Herrmann, 1995), the same is true for $\{\perp_c, \subseteq\}$-atoms taken together (i.e. EMVDs+FDs+INDs in database theory). How can we find complete axioms then?

Axiomatizing $\{\perp_{\mathrm{c}}, \subseteq\}$-atoms

Given a set $\Sigma \cup \{\phi\}$ of $\{\perp_{\mathrm{c}}, \subseteq\}$-atoms, the implication problem $\Sigma \models \phi$ is to decide whether $X \models \phi$ for all teams X such that $X \models \Sigma$. We will present a complete axiomatization of the implication problem for $\{\perp_{\mathrm{c}}, \subseteq\}$-atoms.

Since the implication problem for EMVDs is undecidable (Herrmann, 1995), the same is true for $\{\perp_{\mathrm{c}}, \subseteq\}$-atoms taken together (i.e. EMVDs+FDs+INDs in database theory). How can we find complete axioms then?

- One could axiomatize some larger class of dependencies where these additional dependencies would be needed at intermediate steps of derivations.

Axiomatizing $\{\perp_{\mathrm{c}}, \subseteq\}$-atoms

Given a set $\Sigma \cup \{\phi\}$ of $\{\perp_{\mathrm{c}}, \subseteq\}$-atoms, the implication problem $\Sigma \models \phi$ is to decide whether $X \models \phi$ for all teams X such that $X \models \Sigma$. We will present a complete axiomatization of the implication problem for $\{\perp_{\mathrm{c}}, \subseteq\}$-atoms.

Since the implication problem for EMVDs is undecidable (Herrmann, 1995), the same is true for $\{\perp_{\mathrm{c}}, \subseteq\}$-atoms taken together (i.e. EMVDs+FDs+INDs in database theory). How can we find complete axioms then?

- One could axiomatize some larger class of dependencies where these additional dependencies would be needed at intermediate steps of derivations.
- One could (implicitly) use existential/universal quantification.

Axioms for FDs+INDs

The latter line is used in Mitchell's axiomatization (1983) for FDs and inclusion dependencies (INDs) taken together. The corresponding implication problem is known to be undecidable but the following finite set of rules is still sound and complete.

- Complete axioms for FDs alone (3).
- Complete axioms for INDs alone (3).
- (Identity) From $AB \subseteq CC$ and σ derive τ, where τ is obtained from σ by substituting A for one or more occurrences of B.
- (Pullback) From $UV \subseteq XY$ and $X \rightarrow Y$ derive $U \rightarrow V$, where $|X| = |U|$.
- (Collection) From $UV \subseteq XY$, $UW \subseteq XZ$ and $X \rightarrow Y$ derive $UVW \subseteq XYZ$, where $|X| = |U|$.
- (Attribute Introduction) From $U \subseteq V$ and $V \rightarrow B$ derive $UA \subseteq VB$.

Axioms for FDs+INDs cont.

Note that Attribute Introduction

- from $U \subseteq V$ and $V \to B$ derive $UA \subseteq VB$,

is not sound in the usual sense. For instance the following relation

	U	V	A	B
t_0	0	0	2	1
t_1	0	0	3	1

satisfies $U \subseteq V$ and $V \to B$, but not $UA \subseteq VB$.

Axioms for FDs+INDs cont.

Note that Attribute Introduction

- from $U \subseteq V$ and $V \to B$ derive $UA \subseteq VB$,

is not sound in the usual sense. For instance the following relation

	U	V	A	B
t_0	0	0	1	1
t_1	0	0	1	1

satisfies $U \subseteq V$ and $V \to B$, but not $UA \subseteq VB$.

Soundness is obtained if we assume that the attribute A does not appear earlier in the proof, and no new variables of this kind are allowed to appear at the final step of a derivation. Hence, A should be thought of as implicitly existentially quantified.

Axioms for $\{\perp_{\rm c}, \subseteq\}$-atoms I

In this style we can axiomatize $\{\perp_{\rm c}, \subseteq\}$-atoms also:

1. Reflexivity:
$$\vec{x} \subseteq \vec{x}.$$
2. Projection and Permutation:
$$\text{if } x_1 \ldots x_n \subseteq y_1 \ldots y_n, \text{ then } x_{i_1} \ldots x_{i_k} \subseteq y_{i_1} \ldots y_{i_k},$$
for each sequence $i_1, \ldots, i_k$ of integers from $\{1, \ldots, n\}$.
3. Transitivity:
$$\text{if } \vec{x} \subseteq \vec{y} \wedge \vec{y} \subseteq \vec{z}, \text{ then } \vec{x} \subseteq \vec{y}.$$
4. Identity Rule:
$$\text{if } ab \subseteq cc \wedge \phi, \text{ then } \phi',$$
where ϕ' is obtained from ϕ by replacing any number of occurrences of a by b.

Axioms for $\{\perp_c, \subseteq\}$-atoms II

5. Inclusion Introduction:

$$\text{if } \vec{a} \subseteq \vec{b}, \text{ then } \vec{a}x \subseteq \vec{b}c,$$

   where x is a *new* variable.

6. Start Axiom:

$$\vec{a}\vec{c} \subseteq \vec{a}\vec{x} \wedge \vec{b} \perp_{\vec{a}} \vec{x} \wedge \vec{a}\vec{x} \subseteq \vec{a}\vec{c}$$

   where $\vec{x}$ is a sequence of pairwise distinct *new* variables.

7. Chase Rule:

$$\text{if } \vec{y} \perp_{\vec{x}} \vec{z} \wedge \vec{a}\vec{b} \subseteq \vec{x}\vec{y} \wedge \vec{a}\vec{c} \subseteq \vec{x}\vec{z}, \text{ then } \vec{a}\vec{b}\vec{c} \subseteq \vec{x}\vec{y}\vec{z}.$$

8. Final Rule:

$$\text{if } \vec{a}\vec{c} \subseteq \vec{a}\vec{x} \wedge \vec{b} \perp_{\vec{a}} \vec{x} \wedge \vec{a}\vec{b}\vec{x} \subseteq \vec{a}\vec{b}\vec{c}, \text{ then } \vec{b} \perp_{\vec{a}} \vec{c}.$$

Axioms for $\{\perp_c, \subseteq\}$-atoms III

A deduction from Σ is a sequence of formulas $(\phi_1, \ldots, \phi_n)$ such that:

1. Each ϕ_i is either an element of Σ, an instance of Reflexivity or Start Axiom, or follows from one or more formulas of $\Sigma \cup \{\phi_1, \ldots, \phi_{i-1}\}$ by one of the rules.
2. If ϕ_i is an instance of Start Axiom (or follows from $\Sigma \cup \{\phi_1, \ldots, \phi_{i-1}\}$ by Inclusion Introduction), then the new variables of $\vec{x}$ (or the new variable x) must not appear in $\Sigma \cup \{\phi_1, \ldots, \phi_{i-1}\}$.

Axioms for $\{\perp_c, \subseteq\}$-atoms IV

We say that ϕ is provable from Σ, written $\Sigma \vdash \phi$, if there is a deduction $(\phi_1, \ldots, \phi_n)$ from Σ with $\phi = \phi_n$ and such that no variables in ϕ are new in $\phi_1, \ldots, \phi_n$.

On Soundness: Inclusion Introduction

Definition

$$\text{If } \vec{a} \subseteq \vec{b}, \text{ then } \vec{a}x \subseteq \vec{b}c,$$

where x is a *new* variable.

This rule is sound if we interpret x as existentially quantified. For instance $X = \{s_0, s_1, s_2, s_3\}$ satisfies $a \subseteq b$, and it can be extended to satisfy $ax \subseteq bc$.

	a	b	c	
s_0	2	3	1	
s_1	4	4	7	
s_2	5	2	6	
s_3	3	5	8	

On Soundness: Inclusion Introduction

Definition

$$\text{If } \vec{a} \subseteq \vec{b}, \text{ then } \vec{a}x \subseteq \vec{b}c,$$

where x is a *new* variable.

This rule is sound if we interpret x as existentially quantified. For instance $X = \{s_0, s_1, s_2, s_3\}$ satisfies $a \subseteq b$, and it can be extended to satisfy $ax \subseteq bc$.

	a	b	c	x
s_0	2	3	1	6
s_1	4	4	7	7
s_2	5	2	6	8
s_3	3	5	8	1

On Soundness: Start Axiom

Definition

$$\vec{a}\vec{c} \subseteq \vec{a}\vec{x} \wedge \vec{b} \perp_{\vec{a}} \vec{x} \wedge \vec{a}\vec{x} \subseteq \vec{a}\vec{c}$$

where $\vec{x}$ is a sequence of pairwise distinct *new* variables.

This rule is sound if we interpret $\vec{x}$ as existentially quantified in the (lax) team semantics sense:

Definition (Ex. Quant. in TS)

$\mathcal{M} \models_X \exists v\psi$ if and only if there is a function $F : X \to \mathcal{P}(M)\setminus\{\emptyset\}$ such that $\mathcal{M} \models_{X[F/v]} \psi$, where $X[F/v] = \{s[m/v] : s \in X, m \in F(s)\}$.

On Soundness: Start Axiom cont.

Definition

$$\vec{a}\vec{c} \subseteq \vec{a}\vec{x} \wedge \vec{b} \perp_{\vec{a}} \vec{x} \wedge \vec{a}\vec{x} \subseteq \vec{a}\vec{c}$$

where $\vec{x}$ is a sequence of pairwise distinct *new* variables.

Let X be a team. If we extend each assignment $s \in X$ with a set of values for x as below, then X satisfies $ac \subseteq ax \wedge b \perp_a x \wedge ax \subseteq ac$.

	a	b	c	
	$\vdots$	$\vdots$	$\vdots$	
s	1	2	3	
	$\vdots$	$\vdots$	$\vdots$	

On Soundness: Start Axiom cont.

Definition

$$\vec{a}\vec{c} \subseteq \vec{a}\vec{x} \wedge \vec{b} \perp_{\vec{a}} \vec{x} \wedge \vec{a}\vec{x} \subseteq \vec{a}\vec{c}$$

where $\vec{x}$ is a sequence of pairwise distinct *new* variables.

Let X be a team. If we extend each assignment $s \in X$ with a set of values for x as below, then X satisfies $ac \subseteq ax \wedge b \perp_a x \wedge ax \subseteq ac$.

	a	b	c	x
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
s	1	2	3	$\{s'(c) : s' \in X, s'(a) = 1\}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Soundness: Conclusion

Theorem

Let $\Sigma \cup \{\phi\}$ be a finite set of conditional independence and inclusion atoms. Then $\Sigma \models \phi$ if $\Sigma \vdash \phi$.

Proof.

Idea. Assume that $X \models \Sigma$; we have to show that $X \models \phi$. By the assumption there is a deduction $(\phi_1, \ldots, \phi_n)$ from Σ with $\phi = \phi_n$ and such that no variables in ϕ are new in $\phi_1, \ldots, \phi_n$. By the previous considerations we find a team $X' \models \phi$ which extends X with values for the new variables. Since no new variables appear in ϕ, we obtain that $X \models \phi$.

$\square$

On completeness

Theorem

Let $\Sigma \cup \{\phi\}$ be a finite set of conditional independence and inclusion atoms. Then $\Sigma \vdash \phi$ if $\Sigma \models \phi$.

Proof.

Idea. A chase characterization of the implication problem is presented: for each $\Sigma \cup \{\phi\}$, we inductively construct an infinite graph $G_{\Sigma,\phi}$, consisting of nodes and labeled edges, and such that $\Sigma \models \phi$ exactly when we find a node $v \in G_{\Sigma,\phi}$ which is connected to the initial stage in a certain way. The construction of $G_{\Sigma,\phi}$ can be simulated in our axiom system from which the result will follow. □