INFORMATIONAL REPRESENTABILITY FOR CONTEXTS IN DEDEKIND CATEGORIES

Okuma, Hitomi
Department of Informatics, Kyushu University

MacCaull, Wendy
Department of Mathematics, Statistics and Computer Science, St. Francis Xavier University

Kawahara, Yasuo
Department of Informatics, Kyushu University

INFORMATIONAL REPRESENTABILITY FOR CONTEXTS IN DEDEKIND CATEGORIES

By

Hitomi OKUMA *, Wendy MacCAULL [†] and   Yasuo KAWAHARA[‡]

Abstract

This paper presents a categorical formulation for association rules in information systems. MacCaull developed a tableaux style calculus that is sound and complete for the implication problem for association rules. The proof of the completeness required an informational representability result, to go from frames to contexts with 1. To show the completeness for the implication problem for association rules in our categorical framework, we prove an informational representability result in Dedekind categories.

1.   Introduction

Relational database theory has been studied since Codd (1970) introduced relational database models. Dependency theory is concerned with the general problem of relationships among attributes in database relations. Database dependency theory has been studied with relational methods. Orlowska (1987) proposed a relational formulation of functional, multivalued and other dependencies, and Buszkowski and Orlowska (1998) developed an axiomatic relational calculus for dependency theory. Okuma and Kawahara (2000, 2001) extend some database dependency theory to Dedekind categories. Schmidt and Ströhlein (1993) explained a basic relational feature of functional dependency for relational models of databases. The foundations and recent applications of relational methods in computer science are excellently summarized in Brink et al. (1997).

MacCaull (2000, 2001) investigated a relational formulation for functional and multivalued dependencies and association rules, and proved soundness and completeness for the implication problem of these dependencies with a Rasiowa/Sikorski-style tableaux proof system. MacCaull (2001) focuses on association rules in contexts, which are databases such that all attribute values are either 0 or 1. There the notion of association relation was used to express association rules and sufficient conditions were given to prove an informational representability result used in the proof of the completeness of a tableaux deduction method for the implication problem for association rules.

The aim of this paper is to give a categorical formulation for information systems and to prove the informational representability for contexts in our categorical framework

---

* Department of Informatics, Kyushu University 33 Fukuoka 812-8581 Japan. okuma@i.kyushu-u.ac.jp

† Department of Mathematics, Statistics and Computer Science, St.Francis Xavier University, PO Box 5000, Antigonish, NS, B2G 2W5, Canada. wmaccaul@stfx.ca

‡ Department of Informatics, Kyushu University 33 Fukuoka 812-8581 Japan. kawahara@i.kyushu-u.ac.jp

in order to extend the work on association rules to Dedekind categories.

The remainder of this paper is organized as follows: In Section 2, we first review the notion of information systems, a generalization of database relation to indeterminate databases (of use both in rough set theory and in classical database theory). Then we review the notion of contexts, association relation and association rule in a context with 1.

In Section 3, we briefly review the definition of Dedekind category, a kind of relation category, and list basic properties of relations in Dedekind categories. Next we present the definition and some properties of binomial equivalence relations in order to discuss association rules in Dedekind categories.

In Section 4, we give a definition of information system in a Dedekind category. We show some properties of indiscernibility relations found in MacCaull (2000, 2001) hold in Dedekind categories.

Next we generalize the notions of context with 1, association relation and associa- tion rule in contexts with 1 to Dedekind categories. We show some basic properties of association relations in a Dedekind category, which are generalizations of the results in MacCaull (2001). Finally we define the notion of information frame in a Dedekind cate- gory and prove that the informational representability result found in MacCaull (2001) holds.

2.   Information Systems

In this section, we review the definition of information systems and the formulation of some constraints in information systems.

First we recall a foundation of (binary) relational calculus on sets. A relation R of a set A to a set B, denoted by a half arrow $R : A \rightharpoonup B$, is a subset of the cartesian product $A \times B$. Given relations $R : A \rightharpoonup B$, $S : B \rightharpoonup C$ and $T : A \rightharpoonup C$, the operations of composition, conversion and residue are defined as follows:

$$RS = \{(x,z) \in A \times C \mid \text{for some } y \in B, \ (x,y) \in R \text{ and } (y,z) \in S\},$$

$$R^\sharp = \{(x,y) \in B \times A \mid (y,x) \in R\} \text{ and}$$

$$T \div S = \{(x,y) \in A \times B \mid \text{for all } z \in C, \ (y,z) \in S \text{ implies } (x,z) \in T\}.$$

As relations are subsets of a cartesian product, the inclusion $\subseteq$, the union $\cup$, the inter- section $\cap$ and the complement $^-$ are available as usual.

Now we review the definition of information systems as presented in MacCaull (2000).

An *information system* is a 4-tuple $(X, U, \{V_a : a \in U\}, f)$, where X is a set (of tuples); U is a set (of attributes); for each $a \in U$, V_a is the set of values of an attribute a, and for each $x \in X$ and $a \in U$, $f(x,a) \subseteq V_a$. This is a generalization of database relation to the situation of indeterminate databases, and it is of use both in rough set theory (see Orłowska, 1998) and classical database theory. If $f(x,a) \in V_a$ then this is a database relation with attributes in U.

For each $P \subseteq U$, an *indiscernibility relation* $R_P^{ind} : X \rightharpoonup X$ is defined by

$$R_P^{ind} = \{(x,y) \in X \times X | \forall a \in P, f(x,a) = f(y,a)\}.$$

Using the indiscernibility relations one may express many dependencies in relational databases (see Orłowska, 1987; Buszkowski and Orłowska, 1998; MacCaull, 2000). For example, let P and Q be subsets of U:

a functional dependency $P \to Q$ holds in X
iff $\forall x, y \in X, \{(\forall a \in P, f(x,a) = f(y,a)) \Rightarrow (\forall a \in Q, f(x,a) = f(y,a))\}$
iff $R_P^{ind} \subseteq R_Q^{ind}$.

Given an information system $(X, U, \{V_a : a \in U\}, f)$, the following hold for all subsets P and Q of U: (a) R_P^{ind} is an equivalence relation, (b) $R_P^{ind} \cap R_Q^{ind} = R_{P \cup Q}^{ind}$ and (c) $R_\emptyset^{ind} = X \times X$. The properties (b) and (c) imply that the set of indiscernibility relations is a set of strong relations (see Orłowska, 1998).

Now we recall the definitions of context, association relation and association rule in a context with 1 (MacCaull, 2000, 2001).

A *context* is an information system $(X, U, \{V_a : a \in U\}, f)$ such that all attribute values are either 0 or 1; that is, $f(x,a) \in V_a = \{0,1\}$ for all $x \in X$ and for all $a \in U$. A *context with* 1 is a context with an object 1_U in X satisfying $f(1_U, a) = 1$ for all $a \in U$. An *association rule* is a dependency for contexts with 1 defined as follows: let P and Q be subsets of U, an association rule $P \Rightarrow Q$ holds in X iff $\forall x \in X, f(x,a) = 1$ for all $a \in P$ implies $f(x,a) = 1$ for all $a \in Q$.

An *association relation* $R_P^{as} : X \rightharpoonup X$ in a context with 1 is defined as follows:

$$R_P^{as} = \{(x,y) \in X \times X | \forall a \in P, f(x,a) = 1 \text{ iff } \forall a \in P, f(y,a) = 1\}.$$

A relational formulation for the association rule $P \Rightarrow Q$ may be given as follows:

$$P \Rightarrow Q \text{ iff } R_P^{as} = R_{P \cup Q}^{as}.$$

Given a context with 1, the following hold for all subsets P and Q of U: (a) R_P^{as} is an equivalence relation, (b) $R_P^{as} \cap R_Q^{as} \subseteq R_{P \cup Q}^{as}$ and (c) $R_\emptyset^{as} = X \times X$. The properties (b) and (c) imply that the set of association relations is a set of semistrong relations.

Fact 1 Let $(X, U, \{V_a : a \in U\}, f)$ be an information system and $f_a : X \to \wp(V_a)$ be a function such that $f_a(x) = f(x,a)$ for all $x \in X$ and for all $a \in U$. Then we have $\cap_{a \in P} f_a f_a^\sharp = R_P^{ind}$ for each subset P of U.

PROOF. The result follows from

$(x,y) \in \cap_{a \in P} f_a f_a^\sharp$
iff $\forall a \in P, (x,y) \in f_a f_a^\sharp$
iff $\forall a \in P, \exists v \in \wp(V_a), v = f_a(x) \wedge v = f_a(y)$ { f_a is a function }
iff $\forall a \in P, f(x,a) = f(y,a)$
iff $(x,y) \in R_P^{ind}$.

□

The next fact shows that an association relation can be expressed using indiscernibility relations and residues.

Fact 2 Let $(X, U, \{V_a : a \in U\}, f)$ be a context with 1, $t : \{*\} \to X$ be a function from a singleton set $\{*\}$ to X such that $t(*) = 1_U$. Then $\{(tR_P^{ind})^\sharp \div (tR_P^{ind})^\sharp\} \cap \{(tR_P^{ind})^\sharp \div (tR_P^{ind})^\sharp\}^\sharp = R_P^{as}$ for each subset P of U.

PROOF. First we have

$$(x, y) \in (tR_P^{ind})^\sharp \div (tR_P^{ind})^\sharp$$

iff $\quad (y, *) \in (tR_P^{ind})^\sharp$ implies $(x, *) \in (tR_P^{ind})^\sharp$

iff $\quad (1_U, y) \in R_P^{ind}$ implies $(1_U, x) \in R_P^{ind}$

iff $\quad \forall a \in P : f(1_U, a) = f(y, a)$ implies $\forall a \in P : f(1_U, a) = f(x, a).$

In the same way we have

$$(x, y) \in \{(tR_P^{ind})^\sharp \div (tR_P^{ind})^\sharp\}^\sharp$$

iff $\quad \forall a \in P : f(1_U, a) = f(x, a)$ implies $\forall a \in P : f(1_U, a) = f(y, a).$

Then we conclude

$$(x, y) \in \{(tR_P^{ind})^\sharp \div (tR_P^{ind})^\sharp\} \cap \{(tR_P^{ind})^\sharp \div (tR_P^{ind})^\sharp\}^\sharp$$

iff $\quad \forall a \in P : f(x, a) = f(1_U, a)$ if and only if $\forall a \in P : f(y, a) = f(1_U, a)$

iff $\quad (x, y) \in R_P^{as}.$

$\square$

3. Dedekind Categories

In this section, we first recall the definition of a Dedekind category, a kind of relation category (following Olivier and Serrato, 1980) which is our general framework, and then present some properties of relations in Dedekind categories. The composition operator will bind stronger than all other binary operators.

Throughout this paper, a morphism α from an object A into an object B in a Dedekind category (which will be defined below) will be called a relation, and denoted by a half arrow $\alpha : A \rightharpoonup B$. The composite of a relation $\alpha : A \rightharpoonup B$ followed by a relation $\beta : B \rightharpoonup C$ will be written as $\alpha\beta : A \rightharpoonup C$. We denote the identity relation on an object A by id_A.

DEFINITION 3.1. A *Dedekind category* $\mathcal{D}$ is a category satisfying the following:
D1. [Complete Distributive Lattice] For all pairs of objects A and B the hom-set $\mathcal{D}(A, B)$ consisting of all relations of A into B is a complete distributive lattice with the least relation 0_{AB} and the greatest relation ∇_{AB}. Its algebraic structure will be denoted by

$$\mathcal{D}(A, B) = (\mathcal{D}(A, B), \sqsubseteq, \sqcup, \sqcap, 0_{AB}, \nabla_{AB}).$$

That is, (a) $\sqsubseteq$ is a partial order on $\mathcal{D}(A, B)$, (b) $\forall \alpha \in \mathcal{D}(A, B) :: 0_{AB} \sqsubseteq \alpha \sqsubseteq \nabla_{AB}$, (c) $\sqcup_{j \in J}\alpha_j \sqsubseteq \alpha$ iff $\alpha_j \sqsubseteq \alpha$ for all $j \in J$, (d) $\alpha \sqsubseteq \sqcap_{j \in J}\alpha_j$ iff $\alpha \sqsubseteq \alpha_j$ for all $j \in J$, and (e) $\alpha \sqcap (\sqcup_{j \in J}\alpha_j) = \sqcup_{j \in J}(\alpha \sqcap \alpha_j).$
D2. [Converse] For all objects A and B there is given an operator $\sharp : \mathcal{D}(A, B) \to \mathcal{D}(B, A)$

such that for all relations $\alpha, \alpha' : A \to B$, $\beta : B \to C$, the following laws hold:
(a) $(\alpha\beta)^\sharp = \beta^\sharp\alpha^\sharp$, (b) $(\alpha^\sharp)^\sharp = \alpha$, (c) If $\alpha \sqsubseteq \alpha'$ then $\alpha^\sharp \sqsubseteq \alpha'^\sharp$.
D3. [Dedekind Formula] For all relations $\alpha : A \to B$, $\beta : B \to C$ and $\gamma : A \to C$ the Dedekind formula $\alpha\beta \sqcap \gamma \sqsubseteq \alpha(\beta \sqcap \alpha^\sharp\gamma)$ holds.
D4. [Residues] For all relations $\beta : B \to C$ and $\gamma : A \to C$ the residue (or division or weakest precondition) $\gamma \div \beta : A \to B$ is a relation such that $\alpha\beta \sqsubseteq \gamma$ if and only if $\alpha \sqsubseteq \gamma \div \beta$ for all relations $\alpha : A \to B$. $\qquad\square$

In a Dedekind category $\mathcal{D}$ a function $f : A \to B$ is a relation $f : A \to B$ such that $f^\sharp f \sqsubseteq \mathrm{id}_B$ (univalent) and $\mathrm{id}_A \sqsubseteq ff^\sharp$ (total), and an equivalence relation $\eta : A \to A$ is a relation such that $\mathrm{id}_A \sqsubseteq \eta$ (reflexive), $\eta^\sharp \sqsubseteq \eta$ (symmetric) and $\eta\eta \sqsubseteq \eta$ (transitive).

An object I of a Dedekind category $\mathcal{D}$ is called a *unit* if $0_{II} \neq \mathrm{id}_I = \nabla_{II}$. A unit I is called *strict* if $\nabla_{AI}\nabla_{IA} = \nabla_{AA}$ for all objects A of $\mathcal{D}$. An *I-point* x of A is a function $x : I \to A$.

In the next proposition we collect some basic properties of relations in Dedekind categories. The proof of the next proposition and more details on fundamental properties of relations may be found in Freyd and Scedrov (1990), Schmidt and Ströhlein (1993) or Furusawa and Kahl (1998).

PROPOSITION 3.2. *Let* $\alpha, \alpha' : A \to B$, $\beta, \beta', \beta_j : B \to C$, $\gamma, \gamma', \gamma_j : A \to C$, $\delta : D \to B$ *be relations and* $\zeta : E \to A$ *a function in a Dedekind category* $\mathcal{D}$. *Then the following hold:*

(a) $0_{AB}^\sharp = 0_{BA}$, $\nabla_{AB}^\sharp = \nabla_{BA}$ *and* $\mathrm{id}_A^\sharp = \mathrm{id}_A$.

(b) $(\alpha \sqcup \alpha')^\sharp = \alpha^\sharp \sqcup \alpha'^\sharp$ *and* $(\alpha \sqcap \alpha')^\sharp = \alpha^\sharp \sqcap \alpha'^\sharp$.

(c) $\alpha(\sqcup_{j \in J}\beta_j) = \sqcup_{j \in J}\alpha\beta_j$.

(d) $\alpha(\sqcap_{j \in J}\beta_j) \sqsubseteq \sqcap_{j \in J}\alpha\beta_j$.

(e) *If* $\alpha \sqsubseteq \alpha'$ *and* $\beta \sqsubseteq \beta'$ *then* $\alpha\beta \sqsubseteq \alpha'\beta'$.

(f) *If* α *is univalent then* $\alpha(\sqcap_{j \in J}\beta_j) = \sqcap_{j \in J}\alpha\beta_j$.

(g) $(\gamma \div \beta)\beta \sqsubseteq \gamma$, $\gamma \div \mathrm{id}_C = \gamma$, $\nabla_{AC} \div \beta = \nabla_{AB}$ *and* $\gamma \div 0_{BC} = \nabla_{AB}$.

(h) $(\sqcap_{j \in J}\gamma_j) \div \beta = \sqcap_{j \in J}(\gamma_j \div \beta)$ *and* $\gamma \div (\sqcup_{j \in J}\beta_j) = \sqcap_{j \in J}(\gamma \div \beta_j)$.

(i) $\gamma \div (\delta\beta) = (\gamma \div \beta) \div \delta$.

(j) *If* $\gamma \sqsubseteq \gamma'$ *and* $\beta' \sqsubseteq \beta$ *then* $\gamma \div \beta \sqsubseteq \gamma' \div \beta'$.

(k) $\zeta(\gamma \div \beta) = (\zeta\gamma) \div \beta$.

(l) *If* β *is a function then* $\gamma \div \beta = \gamma\beta^\sharp$. $\qquad\square$

Next we recall the definition of binomial equivalence relations, which generalizes the notion of equivalence relation with at most two equivalence classes.

DEFINITION 3.3. For a relation $\rho : I \to A$ define a *binomial equivalence relation* $\xi(\rho) : A \to A$ as:

$$\xi(\rho) = (\rho^\sharp \div \rho^\sharp) \sqcap (\rho^\sharp \div \rho^\sharp)^\sharp.$$

$\square$

The next lemma collects some properties of binomial equivalence relations. Later on, we will use them to show the properties of association relations in Dedekind categories.

LEMMA 3.4. *Let* ρ, ρ_1, $\rho_2 : I \to A$ *and* $\eta : A \to A$ *be relations and* $t : I \to A$ *an* I-*point of* A *in a Dedekind category* $\mathcal{D}$. *Then the following hold:*

(a) $\xi(\rho) : A \to A$ *is an equivalence relation on* A.

(b) $\xi(\rho_1) \sqcap \xi(\rho_2) \sqsubseteq \xi(\rho_1 \sqcap \rho_2)$.

(c) $\xi(\nabla_{IA}) = \xi(0_{IA}) = \nabla_{AA}$.

(d) *If* η *is an equivalence relation then* $\eta \sqsubseteq \xi(t\eta)$.

(e) *If* η *is reflexive then* $t\,\xi(t\eta) = t\eta$.

PROOF. (a) First reflexivity follows from the next equivalence: $\mathrm{id}_A \sqsubseteq \rho^\sharp \div \rho^\sharp$ iff $\mathrm{id}_A \rho^\sharp \sqsubseteq \rho^\sharp$. $\xi(\rho)$ is clearly symmetric by the definition. By Proposition 3.2(g) we have $(\rho^\sharp \div \rho^\sharp)(\rho^\sharp \div \rho^\sharp) \sqsubseteq \rho^\sharp \div \rho^\sharp$, and hence transitivity follows from

$$
\begin{aligned}
\xi(\rho)\xi(\rho) &= \{(\rho^\sharp \div \rho^\sharp) \sqcap (\rho^\sharp \div \rho^\sharp)^\sharp\}\{(\rho^\sharp \div \rho^\sharp) \sqcap (\rho^\sharp \div \rho^\sharp)^\sharp\} \\
&\sqsubseteq (\rho^\sharp \div \rho^\sharp)(\rho^\sharp \div \rho^\sharp) \sqcap (\rho^\sharp \div \rho^\sharp)^\sharp(\rho^\sharp \div \rho^\sharp)^\sharp \qquad \{ \text{ Proposition 3.2(d) } \} \\
&\sqsubseteq (\rho^\sharp \div \rho^\sharp) \sqcap (\rho^\sharp \div \rho^\sharp)^\sharp \\
&= \xi(\rho).
\end{aligned}
$$

(b) It is enough to show that

$$(\rho_1^\sharp \div \rho_1^\sharp) \sqcap (\rho_2^\sharp \div \rho_2^\sharp) \sqsubseteq (\rho_1 \sqcap \rho_2)^\sharp \div (\rho_1 \sqcap \rho_2)^\sharp.$$

This follows from

$$
\begin{aligned}
&(\rho_1^\sharp \div \rho_1^\sharp) \sqcap (\rho_2^\sharp \div \rho_2^\sharp) \\
\sqsubseteq\ & \{\rho_1^\sharp \div (\rho_1^\sharp \sqcap \rho_2^\sharp)\} \sqcap \{\rho_2^\sharp \div (\rho_1^\sharp \sqcap \rho_2^\sharp)\} \quad \{ \text{ Proposition 3.2(j) } \} \\
=\ & (\rho_1^\sharp \sqcap \rho_2^\sharp) \div (\rho_1^\sharp \sqcap \rho_2^\sharp) \qquad\qquad\quad\ \{ \text{ Proposition 3.2(h) } \} \\
=\ & (\rho_1 \sqcap \rho_2)^\sharp \div (\rho_1 \sqcap \rho_2)^\sharp. \qquad\qquad\ \ \{ \text{ Proposition 3.2(b) } \}
\end{aligned}
$$

(c) By Proposition 3.2(g), (a) we have $\xi(\nabla_{IA}) = (\nabla_{IA}^\sharp \div \nabla_{IA}^\sharp) \sqcap (\nabla_{IA}^\sharp \div \nabla_{IA}^\sharp)^\sharp = \nabla_{AA}$, and $\xi(0_{IA}) = (0_{IA}^\sharp \div 0_{IA}^\sharp) \sqcap (0_{IA}^\sharp \div 0_{IA}^\sharp)^\sharp = \nabla_{AA}$.

(d) Assume that η is an equivalence relation. Then we have $\eta(t\eta)^\sharp = (t\eta)^\sharp$ and so $\eta \sqsubseteq (t\eta)^\sharp \div (t\eta)^\sharp$. Hence $\eta \sqsubseteq \xi(t\eta)$.

(e) Assume that η is reflexive. We first get $\mathrm{id}_A \sqsubseteq \eta^\sharp$ since η is reflexive, and so $\mathrm{id}_I \sqsubseteq tt^\sharp \sqsubseteq t\eta^\sharp t^\sharp = t(t\eta)^\sharp$ since t is total. Then we have

$$
\begin{aligned}
t\{(t\eta)^\sharp \div (t\eta)^\sharp\} &= \{t(t\eta)^\sharp\} \div (t\eta)^\sharp \quad \{ \text{t is function and Proposition 3.2(k) } \} \\
&= \mathrm{id}_I \div (t\eta)^\sharp \qquad\quad\ \{ \text{$t(t\eta)^\sharp = \mathrm{id}_I$ } \} \\
&= \nabla_{IA}, \qquad\qquad\qquad \{ \text{$\mathrm{id}_I = \nabla_{II}$ and Proposition 3.2(g) } \}
\end{aligned}
$$

and

$$\{(t\eta)^{\sharp} \div (t\eta)^{\sharp}\}t^{\sharp} = \{(t\eta)^{\sharp} \div (t\eta)^{\sharp}\} \div t \quad \{ \ t \text{ is function and Proposition 3.2(l) } \}$$
$$= (t\eta)^{\sharp} \div \{t(t\eta)^{\sharp}\} \quad \{ \text{ Proposition 3.2(i) } \}$$
$$= (t\eta)^{\sharp} \div \mathrm{id}_I \quad \{t(t\eta)^{\sharp} = \mathrm{id}_I\}$$
$$= (t\eta)^{\sharp}. \quad \{ \text{ Proposition 3.2(g) } \}$$

By the univalency of t we have the assertion as follows

$$t\,\xi(t\eta) = t[\{(t\eta)^{\sharp} \div (t\eta)^{\sharp}\} \sqcap \{(t\eta)^{\sharp} \div (t\eta)^{\sharp}\}^{\sharp}]$$
$$= t\{(t\eta)^{\sharp} \div (t\eta)^{\sharp}\} \sqcap t\{(t\eta)^{\sharp} \div (t\eta)^{\sharp}\}^{\sharp} \quad \{ \text{ Proposition 3.2(f) } \}$$
$$= \nabla_{IA} \sqcap \{(t\eta)^{\sharp}\}^{\sharp}$$
$$= t\eta.$$

$\square$

We define a notion of complementary hom-sets and Dedekind categories.

DEFINITION 3.5. A hom-set $\mathcal{D}(A,B)$ is *complemented*, if each relation $\alpha : A \rightharpoondown B$ in $\mathcal{D}$ has a complement relation $\alpha^- : A \rightharpoondown B$ such that $\alpha \sqcap \alpha^- = 0_{AB}$ and $\alpha \sqcup \alpha^- = \nabla_{AB}$. A Dedekind category $\mathcal{D}$ will be called complemented if for each object A, the hom-set $\mathcal{D}(A,A)$ is complemented. $\square$

If each hom-set $\mathcal{D}(I,A)$ is complemented, then $\rho^{\sharp} \div \rho^{\sharp} = (\rho^{\sharp-}\rho)^-$ for each $\rho : I \rightharpoondown A$, and so $\xi(\rho) = \rho^{\sharp}\rho \sqcup \rho^{-\sharp}\rho^-$ if I is a strict unit. It follows from

$$(\rho^{\sharp}\rho \sqcup \rho^{-\sharp}\rho^-) \sqcap \xi(\rho)^-$$
$$= (\rho^{\sharp}\rho \sqcup \rho^{-\sharp}\rho^-) \sqcap (\rho^{\sharp-}\rho \sqcup \rho^{\sharp}\rho^-)$$
$$= (\rho^{\sharp}\rho \sqcap \rho^{\sharp-}\rho) \sqcup (\rho^{\sharp}\rho \sqcap \rho^{\sharp}\rho^-) \sqcup (\rho^{-\sharp}\rho^- \sqcap \rho^{\sharp-}\rho) \sqcup (\rho^{-\sharp}\rho^- \sqcap \rho^{\sharp}\rho^-)$$
$$= (\rho^{\sharp} \sqcap \rho^{\sharp-})\rho \sqcup \rho^{\sharp}(\rho \sqcap \rho^-) \sqcup \rho^{-\sharp}(\rho^- \sqcap \rho) \sqcup (\rho^{-\sharp} \sqcap \rho^{\sharp})\rho^-$$
$$\quad \{ \text{ Proposition 3.2(f) } \}$$
$$= 0_{AA}$$

and

$$(\rho^{\sharp}\rho \sqcup \rho^{-\sharp}\rho^-) \sqcup \xi(\rho)^- = (\rho^{\sharp}\rho \sqcup \rho^{-\sharp}\rho^-) \sqcup (\rho^{\sharp-}\rho \sqcup \rho^{\sharp}\rho^-)$$
$$= (\rho^{\sharp} \sqcap \rho^{\sharp-})\rho \sqcup (\rho^{-\sharp} \sqcap \rho^{\sharp})\rho^-$$
$$= \nabla_{AI}\rho \sqcup \nabla_{AI}\rho^-$$
$$= \nabla_{AI}(\rho \sqcup \rho^-)$$
$$= \nabla_{AI}\nabla_{IA}$$
$$= \nabla_{AA}. \quad \{ \ I \text{ is a strict unit } \}$$

4. Association Rules in Dedekind Categories

In this section, we introduce a definition of information system in a Dedekind category, and we define the categorical notion of association relations and association rules for information systems in Dedekind categories.

DEFINITION 4.1. Let U be a set (of attributes). An *information system* in a Dedekind category $\mathcal{D}$ is an U-indexed set $\{f_a : X \to V_a \mid a \in U\}$ of functions f_a from an object X into an object V_a in $\mathcal{D}$. Then for each subset P of U define a relation $\eta[P] : X \to X$ by $\eta[P] = \sqcap_{a \in P} f_a f_a^\sharp$. When P is an empty set, we set $\eta[\emptyset] = \nabla_{XX}$. □

The basic properties of the indiscernibility relation found in MacCaull (2001) can be also shown in information systems in Dedekind categories.

LEMMA 4.2. *Given an information system* $\{f_a : X \to V_a \mid a \in U\}$ *in a Dedekind category* $\mathcal{D}$, *the following hold for all subsets* P *and* Q *of* U:

(a) $\eta[P]$ *is an equivalence relation on* X.

(b) $\eta[P \cup Q] = \eta[P] \sqcap \eta[Q]$.

(c) *If* $P \supseteq Q$ *then* $\eta[P] \sqsubseteq \eta[Q]$.

(d) $\eta[P]\eta[Q] \sqsubseteq \eta[P \cap Q]$.

PROOF. (a) Reflexivity follows from $\mathrm{id}_X \sqsubseteq \sqcap_{a \in P} f_a f_a^\sharp = \eta[P]$, since f_a is total for each $a \in P$. We have $\eta[P]^\sharp = (\sqcap_{a \in P} f_a f_a^\sharp)^\sharp = \sqcap_{a \in P} f_a f_a^\sharp = \eta[P]$, and hence $\eta[P]$ is symmetric. Transitivity follows from

$$
\begin{aligned}
\eta[P]\eta[P] &= (\sqcap_{a \in P} f_a f_a^\sharp)(\sqcap_{a \in P} f_a f_a^\sharp) \\
&\sqsubseteq \sqcap_{a \in P} f_a f_a^\sharp f_a f_a^\sharp \qquad \{ \text{ Proposition 3.2(d) } \} \\
&\sqsubseteq \sqcap_{a \in P} f_a f_a^\sharp \qquad \qquad \{ f_a^\sharp f_a \sqsubseteq \mathrm{id}_{V_a} \} \\
&= \eta[P].
\end{aligned}
$$

(b) It follows from

$$
\begin{aligned}
\eta[P \cup Q] &= \sqcap_{a \in P \cup Q} f_a f_a^\sharp \\
&= (\sqcap_{a \in P} f_a f_a^\sharp) \sqcap (\sqcap_{a \in Q} f_a f_a^\sharp) \\
&= \eta[P] \sqcap \eta[Q].
\end{aligned}
$$

(c) Assume that $P \supseteq Q$. Then we have $\eta[P] = \eta[P \cup Q] = \eta[P] \sqcap \eta[Q] \sqsubseteq \eta[Q]$ by (b).

(d) It follows from

$$
\begin{aligned}
\eta[P]\eta[Q] &= (\sqcap_{a \in P} f_a f_a^\sharp)(\sqcap_{a \in Q} f_a f_a^\sharp) \\
&\sqsubseteq \sqcap_{a \in P \cap Q} f_a f_a^\sharp f_a f_a^\sharp \qquad \{ \text{ Proposition 3.2(d) } \} \\
&\sqsubseteq \sqcap_{a \in P \cap Q} f_a f_a^\sharp \qquad \qquad \{ f_a^\sharp f_a \sqsubseteq \mathrm{id}_{V_a} \} \\
&= \eta[P \cap Q].
\end{aligned}
$$

□

Throughout the rest of this section we assume $\mathcal{D}$ is a Dedekind category with a unit I and $\{f_a : X \to V_a \mid a \in U\}$ is an information system in $\mathcal{D}$.

For an I-point $t : I \to X$, corresponding to 1_U in the context with 1, we call the binomial equivalence relation $\xi(t\eta[P]) : X \to X$ as an *association relation* in $\mathcal{D}$.

We show the properties of association relations found in MacCaull (2001) in Dedekind categories.

LEMMA 4.3. *Let* $\{f_a : X \to V_a \mid a \in U\}$ *be an information system with a designated* I-*point* $t : I \to X$ *in* $\mathcal{D}$*. Then the following hold for all subsets* P *and* Q *of* U:

(a) $\xi(t\eta[P])$ *is an equivalence relation on* X.

(b) $\xi(t\eta[P]) \sqcap \xi(t\eta[Q]) \sqsubseteq \xi(t\eta[P \cup Q])$.

(c) $\xi(t\eta[\emptyset]) = \nabla_{XX}$.

PROOF. (a) The claim is immediate from Lemma 3.4(a).
(b) Since t is univalent, we have $t\eta[P \cup Q] = t(\eta[P] \sqcap \eta[Q]) = t\eta[P] \sqcap t\eta[Q]$ by Lemma 4.2(b) and Proposition 3.2(f). Hence we have $\xi(t\eta[P]) \sqcap \xi(t\eta[Q]) \sqsubseteq \xi(t\eta[P] \sqcap t\eta[Q]) = \xi(t\eta[P \cup Q])$ by Lemma 3.4(b).
(c) By the definition of $\eta[\emptyset]$ we have $t\eta[\emptyset] = t\nabla_{XX} = \nabla_{IX}$ since t is total, and so $\xi(t\eta[\emptyset]) = \nabla_{XX}$ by Lemma 3.4(c).
$\square$

We may express association rules in Dedekind categories as follows:

DEFINITION 4.4. Let $\{f_a : X \to V_a \mid a \in U\}$ be an information system with a designated I-point $t : I \to X$ in $\mathcal{D}$. An *association rule* is a formal expression of the form $P \Rightarrow Q$, where P and Q are subsets of U. We say that a association rule $P \Rightarrow Q$ holds in the information system if and only if $\xi(t\eta[P]) \sqsubseteq \xi(t\eta[P \cup Q])$. $\square$

The following lemma presents a simpler condition which is equivalent to the condition of association rule in the above definition.

LEMMA 4.5. *Let* $\{f_a : X \to V_a \mid a \in U\}$ *be an information system with a designated* I-*point* $t : I \to X$ *in* $\mathcal{D}$*. Then the following statements are equivalent for all subsets* P *and* Q *of* U:

(a) $\xi(t\eta[P]) \sqsubseteq \xi(t\eta[P \cup Q])$.

(b) $t\eta[P] \sqsubseteq t\eta[Q]$.

(c) $\xi(t\eta[P]) = \xi(t\eta[P \cup Q])$.

PROOF. First we show that (a) implies (b). Suppose $\xi(t\eta[P]) \sqsubseteq \xi(t\eta[P \cup Q])$. Then we have $t\eta[P] = t\,\xi(t\eta[P]) \sqsubseteq t\,\xi(t\eta[P \cup Q]) = t\eta[P \cup Q] \sqsubseteq t\eta[Q]$ by Lemma 3.4(e) and Lemma 4.2(c). Next we show that (b) implies (c). Suppose $t\eta[P] \sqsubseteq t\eta[Q]$. Then $\xi(t\eta[P]) = \xi(t\eta[P \cup Q])$ is obvious from the fact that $t\eta[P] \sqsubseteq t\eta[Q]$ iff $t\eta[P] = t\eta[P \cup Q]$. Finally the claim (c) implies (a) is obvious. $\square$

Following the last lemma, the validity of the Armstrong axioms for association rules can be proved easily by using properties of $\eta[P]$.

Fact 3 Let $\{f_a : X \to V_a \mid a \in U\}$ be an information system with a designated I-point $t : I \to X$ in $\mathcal{D}$. The following hold for all subsets P, Q and S of U:
(a) If $P \supseteq Q$ holds then $P \Rightarrow Q$ holds.
(b) If $P \Rightarrow Q$ holds then $P \cup S \Rightarrow Q \cup S$ holds.
(c) If $P \Rightarrow Q$ and $Q \Rightarrow S$ hold then $P \Rightarrow S$ holds.

PROOF. (a) By Lemma 4.2(c) we have $t\eta[P] \sqsubseteq t\eta[Q]$, which is equivalent to $\xi(t\eta[P]) \sqsubseteq \xi(t\eta[P \cup Q])$.

(b) Assume that $t\eta[P] \sqsubseteq t\eta[Q]$. Then $t\eta[P \cup S] = t\eta[P] \sqcap t\eta[S] \sqsubseteq t\eta[Q] \sqcap t\eta[S] = t\eta[Q \cup S]$ by Lemma 4.2(b) and Proposition 3.2(f).

(c) Assume that $t\eta[P] \sqsubseteq t\eta[Q]$ and $t\eta[Q] \sqsubseteq t\eta[S]$. Then $t\eta[P] \sqsubseteq t\eta[S]$.   □

The implication problem for association rules in contexts with 1 is the following: The formula $\bigcap_i P_i \Rightarrow Q_i \supset P \Rightarrow Q$ is provable iff for all contexts with 1 if the associations $P_i \Rightarrow Q_i$ hold, for all i, then the association $P \Rightarrow Q$ holds.

MacCaull (2000, 2001) gives sufficient conditions on semistrong equivalence relations to develop a relational proof system that is sound and complete for the implication problem for association rules in contexts with 1. In fact it is shown that (1) if the tableau for the formula $\bigcap_i P_i \Rightarrow Q_i \supset P \Rightarrow Q$ closes, then if for all i, the associations $P_i \Rightarrow Q_i$ hold in a context with 1 then $P \Rightarrow Q$ holds in that context and (2) if the tableau is open, then one can construct a context with 1 where for all i, the associations $P_i \Rightarrow Q_i$ hold but the association $P \Rightarrow Q$ does not hold. The proof of (2) required an informational representability result, to go from information frames to information systems that are contexts with 1 (the completeness of the tableaux system is frame completeness).

In order to show the completeness for the implication problem for association rules in Dedekind categories, we extend the informational representability result found in MacCaull (2001) to Dedekind categories. We first give a definition of information frames in Dedekind categories.

DEFINITION 4.6. An *information frame* in $\mathcal{D}$ is a structure of the form $\mathbf{K} = (X, U, \varepsilon)$, where X is an object in $\mathcal{D}$, U is a finite set and ε is a map $\varepsilon : \wp(U) \to \mathcal{D}(X, X)$ assigning a relation $\varepsilon[P]$ on X to each subset P of U. If each $\varepsilon[P]$ is an equivalence relation, then we say $\mathbf{K}$ is a frame with equivalence relations. The frame is called *semistrong* if and only if for all subsets P and Q of U, $\varepsilon[P] \sqcap \varepsilon[Q] \sqsubseteq \varepsilon[P \cup Q]$ and $\varepsilon[\emptyset] = \nabla_{XX}$.   □

We conclude this section by establishing the theorem on the informational representability for contexts in a Dedekind category.

THEOREM 4.7. *(Informational representability) Let* $\mathbf{K} = (X, U, \varepsilon)$ *be a semistrong information frame with equivalence relations having a designated I-point* $t : I \to X$ *in* $\mathcal{D}$. *Suppose for all subsets* P *of* U, *the relations* $\varepsilon[P]$ *satisfy the following axioms:*

(As1) $\xi(t\varepsilon[P]) \sqsubseteq \varepsilon[P]$.

(As2) *If* $P \supseteq Q$ *then* $t\varepsilon[P] \sqsubseteq t\varepsilon[Q]$.

(Equiv) *For each equivalence relation* $\theta : X \to X$, *there is an object* X/θ *in* $\mathcal{D}$ *together with a surjection* $q : X \twoheadrightarrow X/\theta$ *such that* $\theta = qq^\sharp$ *and* $q^\sharp q = \mathrm{id}_{X/\theta}$.

Then there is an information system with a designated I-point t *in* $\mathcal{D}$ *such that* $\xi(t\eta[P]) = \varepsilon[P]$.

PROOF. Given a semistrong information frame (X, U, ε) with a designated I-point $t : I \to X$ and equivalence relations that satisfy (As1), (As2) and (Equiv), we construct an information system with t, as follows:

By (Equiv) there exists a surjection $q_a : X \twoheadrightarrow X/\varepsilon[\{a\}]$ such that $q_a q_a^\sharp = \varepsilon[\{a\}]$ and $q_a^\sharp q_a = \mathrm{id}_{X/\varepsilon[\{a\}]}$ for each a in U. Set $V_a = X/\varepsilon[\{a\}]$ for each a in U. Then

$\{q_a : X \to V_a \mid a \in U\}$ is a information system in $\mathcal{D}$ and $\eta[P] = \sqcap_{a \in P} q_a q_a^\sharp$ for each subset P of U.

Then we need to show that for all subsets P of U, $\varepsilon[P] = \xi(t\eta[P])$. By (As1) we have $\varepsilon[P] = \xi(t\varepsilon[P])$ since $\varepsilon[P] \sqsubseteq \xi(t\varepsilon[P])$ always holds by Lemma 3.4(d), so it is enough to show $t\eta[P] = t\varepsilon[P]$ for each subset P of U. First we have

$$
\begin{aligned}
\eta[P] &= \sqcap_{a \in P} q_a q_a^\sharp \\
&= \sqcap_{a \in P} \varepsilon[\{a\}] \quad \{ q_a q_a^\sharp = \varepsilon[\{a\}] \} \\
&\sqsubseteq \varepsilon[\cup_{a \in P} \{a\}] \quad \{ \mathbf{K} \text{ is semistrong} \} \\
&= \varepsilon[P]
\end{aligned}
$$

and so $t\eta[P] \sqsubseteq t\varepsilon[P]$. Next we show the reverse containment. We have for all a in P, $t\varepsilon[P] \sqsubseteq t\varepsilon[\{a\}]$ by (As2). Hence $t\varepsilon[P] \sqsubseteq \sqcap_{a \in P} t\varepsilon[\{a\}] = t(\sqcap_{a \in P} \varepsilon[\{a\}]) = t\eta[P]$ by Proposition 3.2(f). We conclude $t\eta[P] = t\varepsilon[P]$. □

5. Conclusion and Future work

In this paper we present a categorical formulation of indiscernibility relations and association relations and show that properties of those relations hold in Dedekind categories. We also prove the categorical analogue of the informational representability result, necessary for completeness of the implication problem.

Much of database dependency theory has been extended to Dedekind categories (Okuma and Kawahara, 2000, 2001). The question arises whether we can extend the work on association rules to Dedekind categories. There are two different approaches one could take:

The first approach is to develop a tableaux-style of deduction for complemented Dedekind categories (in the sense of Definition 3.5). In fact, since the hom-sets are complete distributive lattices, most of the core relational deduction rules are admissible (see, for example, Orłowska, 1997). The condition that for any $\alpha : X \rightharpoonup X$, $\alpha \sqcup \alpha^- = \nabla_{XX}$ is required so that we may call sequences containing the subsequence α, α^- axiomatic; that is this condition gives a branch closing condition for the tableau. It is of interest to see if this approach gives us tableaux style proofs for implications of functional or multi-valued dependencies or association rules in complemented Dedekind categories. It is also of interest to investigate the possibility of a tableaux style of deduction appropriate for other dependencies for complemented Dedekind categories. With such a development deductions in complemented Dedekind categories can be automated in the same way that relational deduction systems are automated (see Little et al., to appear).

The second approach is to give an axiom system of the style in Okuma and Kawahara (2000, 2001) for association rules in Dedekind categories. This involves designing deduction rules corresponding to the properties of semistrong relations and the properties (As1), (As2) and (Equiv) in Theorem 4.7.

References

Brink, C., Kahl, W. and Schmidt, G. (eds.) (1997). *Relational methods in computer science. Advances in Computing Science*, Springer, Wien, New York.

Beeri, C., Fagin, R. and Howard, J.H. (1997). A complete axiomatization for functional and multivalued dependencies in database relations, *Proc. ACM SIGMOD Internat. Conf. on Management of Data, Toronto*, 47-61.

Buszkowski, W. and Orłowska, E. (1998). Indiscernibility-based formalization of dependencies in information systems, in Orłowska (1998), 293-315.

Codd, E.F. (1970). A relational model of data for large shared data banks, *CACM* **13**(6), 377-387.

Freyd, P. and Scedrov, A. (1990). *Categories, Allegories*, North-Holland, Amsterdam.

Furusawa, H. and Kahl, W. (1998). A Study on Symmetric Quotients, Technical Report 1998-06, Fakultät für Informatik, Universität der Bundeswehr München.

Little, R., MacCaull, W. and Spencer, B. (to appear). ReVAT - Relational Validation by Analytic Tableaux, *St. Francis Xavier University Technical Report*.

MacCaull, W. (2000). A proof system for dependencies for information relations, *Fundamenta Informaticae*, **42**, 1-27.

MacCaull, W. (2001). A tableaux procedure for the implication problem for association rules, in: Orłowska, E. and Szalas, A. (eds.) *Relational Methods for Computer Science Applications*, Springer-Physica Verlag, 77-95.

Okuma, H. and Kawahara, Y. (2000). Relational aspects of relational database dependencies, *Bulletin of Informatics and Cybernetics*, **32**(2), 91-104.

Okuma, H. and Kawahara, Y. (2001). A completeness theorem for extended order dependencies on relational attribute models in Dedekind categories, *Proc. RelMiCS'6 and TARSKI in Oisterwijk*, the Netherlands, October 16-21, 171-184.

Olivier, J.P. and Serrato, D. (1980). Catégories de Dedekind. Morphisms dans les catégories de Schröder, *C.R. Acad. Sci. Paris*, **290**, 939-941.

Orłowska, E. (1987). Algebraic approach to database constraints, *Fundamenta Informaticae*, **10**, 57-68.

Orłowska, E. (1997). Relational formalisation of nonclassical logics, in Brink et al. (1997), 90-105.

Orłowska, E. (1998). *Incomplete information : Rough set analysis*, Physica Verlag.

Schmidt, G. and Ströhlein, T. (1993). *Relations and graphs – Discrete Mathematics for Computer Scientists –*, Springer-Verlag, Berlin.