

Chapter 9

Paraconsistency, Chellas's Conditional Logics, and Association Rules

Tetsuya Murai, Yasuo Kudo and Seiki Akama

Dedicated to Jair Minoro Abe for his 60th birthday

Abstract Paraconsistency and its dual paracompleteness are now counted as key concepts in intelligent decision systems because so much inconsistent and incomplete information can be found around us. In this paper, a framework of conditional models for conditional logic and their measure-based extensions are introduced in order to represent association rules in a logical way. Then paracomplete and paraconsistent aspects of conditionals are examined in the framework. Finally we apply conditionals into the definition of association rules in data mining with confidence and consider their extension to the case of Dempster-Shafer theory of evidence serving double-indexed confidence.

Keywords Paraconsistency · Paracompleteness · Conditional logics · Measure-based semantics · Association rules

T. Murai (✉)
Chitose Institute of Science and Technology, Chitose 066-8655, Japan
e-mail: t-murai@photon.chitose.ac.jp

Y. Kudo
Muroran Institute of Technology, Muroran, Japan
e-mail: kudo@csse.muroran-it.ac.jp

S. Akama
C-Republic, 1-20-1 Higashi-Yurigaoka, Asao-ku, Kawasaki 215-0012, Japan
e-mail: akama@jcom.home.ne.jp

9.1 Introduction

The authors have tried to give a kind of logical foundation to data mining. Murai et al. [15–17] tried to present a logical formulation of association rules [1–3] using Chellas’s conditional models for conditional logics [7] and their measure-based extensions (cf. [12–14]). Akama and Abe [6] proposed a comprehensive idea of paraconsistent logic databases as data warehouse based on paraconsistent and annotated logics [4, 5, 8].

In our opinion, paraconsistency and its dual paracompleteness become key concepts in future development of intelligent decision systems because nowadays there are so much inconsistent and incomplete information around us. In classical logic, inconsistency means triviality in the sense that all sentences become theorems. Paraconsistency means inconsistency but non-triviality. Thus we need new kinds of logic like paraconsistent and annotated logics [4, 5, 8]. Paracompleteness is the dual concept of paraconsistency where the excluded middle is not true.

In this paper, we put association rules in a framework of conditional models [7] and their measure-based extensions (cf. [12–14]) and examine their paracomplete and paraconsistent aspects in the framework.

Then we notice that the standard confidence [1] is nothing but a conditional probability where even weights are a priori assigned to each transaction that contains the items in question at the same time. All of such transactions, however, do not necessarily give us such evidence because some co-occurrences might be contingent. For describing such cases we further introduce double-indexed confidence based on Dempster-Shafer theory [19].

9.2 Chellas’s Conditional Models and Their Measure-Based Extensions for Conditional Logics

9.2.1 Standard and Minimal Conditional Models

Given a finite set \mathcal{P} of items as *atomic sentences*, a language $\mathcal{L}_{\text{CL}}(\mathcal{P})$ for conditional logic is formed from \mathcal{P} as the set of sentences closed under the usual propositional operators such as \top , \perp , \neg , \wedge , \vee , \rightarrow , and \leftrightarrow as well as $\Box\rightarrow$ and $\Diamond\rightarrow$ ¹ (two kinds of conditionals) in the following usual way.

1. If $x \in \mathcal{P}$ then $x \in \mathcal{L}_{\text{CL}}(\mathcal{P})$.
2. $\top, \perp \in \mathcal{L}_{\text{CL}}(\mathcal{P})$.
3. If $p \in \mathcal{L}_{\text{CL}}(\mathcal{P})$ then $\neg p \in \mathcal{L}_{\text{CL}}(\mathcal{P})$.
4. If $p, q \in \mathcal{L}_{\text{CL}}(\mathcal{P})$ then $p \wedge q, p \vee q, p \rightarrow q, p \leftrightarrow q, p\Box\rightarrow q, p\Diamond\rightarrow q \in \mathcal{L}_{\text{CL}}(\mathcal{P})$.

¹In [7], Chellas used only $\Box\rightarrow$. The latter connective $\Diamond\rightarrow$ follows Lewis [11].

Chellas [7] introduces two kind of models called standard and minimal. Their relationship is similar to Kripke and Scott-Montague models for modal logics.

Definition 9.1 (Chellas [7], p. 268) A *standard conditional model* \mathcal{M}_{CL} for conditional logic is a structure

$$\langle W, f, v \rangle,$$

where W is a non-empty set of possible worlds, v is a truth-assignment function

$$v : \mathcal{P} \times W \rightarrow \{0, 1\},$$

and f is a function

$$f : W \times 2^W \rightarrow 2^W. \blacksquare$$

The truth conditions for $\Box \rightarrow$ and $\Diamond \rightarrow$ in standard conditional models are given by

1. $\mathcal{M}_{\text{CL}}, w \models p \Box \rightarrow q \stackrel{\text{def}}{\iff} f(w, \|p\|^{\mathcal{M}_{\text{CL}}}) \subseteq \|q\|^{\mathcal{M}_{\text{CL}}}$,
2. $\mathcal{M}_{\text{CL}}, w \models p \Diamond \rightarrow q \stackrel{\text{def}}{\iff} f(w, \|p\|^{\mathcal{M}_{\text{CL}}}) \cap \|q\|^{\mathcal{M}_{\text{CL}}} \neq \emptyset$,

where $\|p\|^{\mathcal{M}_{\text{CL}}} = \{w \in W \mid \mathcal{M}_{\text{CL}}, w \models p\}$. Thus we have the following relationship between the two kind conditionals:

$$p \Box \rightarrow q \leftrightarrow \neg(p \Diamond \rightarrow \neg q).$$

The function f can be regarded as a kind of selection function. That is, $p \Box \rightarrow q$ is true at a world w when q is true at any world selected by f with respect p and w . Similarly, $p \Diamond \rightarrow q$ is true at a world w when q is true at least at one of the worlds selected by f with respect p and w .

A *minimal conditional model* is a Scott-Montague-like extension of standard conditional model [7].

Definition 9.2 (Chellas [7], p. 270) A *minimal conditional model* \mathcal{M}_{CL} for conditional logic is a structure

$$\langle W, g, v \rangle,$$

where W and v are the same ones as in the standard conditional models. The difference is the second term

$$g : W \times 2^W \rightarrow 2^{2^W}. \blacksquare$$

The truth conditions for $\Box \rightarrow$ and $\Diamond \rightarrow$ in a minimal conditional model are given by

1. $\mathcal{M}_{\text{CL}}, w \models p \Box \rightarrow q \stackrel{\text{def}}{\iff} \|q\|^{\mathcal{M}_{\text{CL}}} \in g(w, \|p\|^{\mathcal{M}_{\text{CL}}})$,
2. $\mathcal{M}_{\text{CL}}, w \models p \Diamond \rightarrow q \stackrel{\text{def}}{\iff} (\|q\|^{\mathcal{M}_{\text{CL}}})^c \notin g(w, \|p\|^{\mathcal{M}_{\text{CL}}})$,

Thus we have also the following relationship:

$$p \Box \rightarrow q \leftrightarrow \neg(p \Diamond \rightarrow \neg q).$$

Note that, if the function g satisfies the following condition

$$X \in g(w, \|p\|^{\mathcal{M}_{\text{CL}}}) \Leftrightarrow \cap g(w, \|p\|^{\mathcal{M}_{\text{CL}}}) \subseteq X$$

for every world w and every sentence p , then, by defining

$$f_g(w, \|p\|^{\mathcal{M}_{\text{CL}}}) \stackrel{\text{def}}{=} \cap g(w, \|p\|^{\mathcal{M}_{\text{CL}}}),$$

we have the standard conditional model $\langle W, f_g, v \rangle$ that is equivalent to the original minimal model.

9.2.2 Measure-Based Extensions

Next we introduce measure-based extensions of the previous minimal conditional models. Such extensions are models for graded conditional logics.

Given a finite set \mathcal{P} of items as *atomic sentences*, a language $\mathcal{L}_{\text{gCL}}(\mathcal{P})$ for graded conditional logic is formed from \mathcal{P} as the set of sentences closed under the usual propositional operators such as $\top, \perp, \neg, \wedge, \vee, \rightarrow$, and \leftrightarrow as well as $\Box \rightarrow_k$ and $\Diamond \rightarrow_k$ (*graded conditionals*) for $0 < k \leq 1$ in the usual way.

1. If $x \in \mathcal{P}$ then $x \in \mathcal{L}_{\text{gCL}}(\mathcal{P})$.
2. $\top, \perp \in \mathcal{L}_{\text{gCL}}(\mathcal{P})$.
3. If $p \in \mathcal{L}_{\text{gCL}}(\mathcal{P})$ then $\neg p \in \mathcal{L}_{\text{gCL}}(\mathcal{P})$.
4. If $p, q \in \mathcal{L}_{\text{gCL}}(\mathcal{P})$ then $p \wedge q, p \vee q, p \rightarrow q, p \leftrightarrow q \in \mathcal{L}_{\text{gCL}}(\mathcal{P})$,
5. If $[p, q \in \mathcal{L}_{\text{gCL}}(\mathcal{P})$ and $0 < k \leq 1]$ then $p \Box \rightarrow_k q, p \Diamond \rightarrow_k q \in \mathcal{L}_{\text{gCL}}(\mathcal{P})$.

A graded conditional model is defined as a family of minimal conditional model (cf. Chellas [7]):

Definition 9.3 Given a fuzzy measure

$$m : 2^W \times 2^W \rightarrow [0, 1],$$

a *measure-based conditional model* $\mathcal{M}_{\text{gCL}}^m$ for graded conditional logic is a structure

$$\langle W, \{g_k\}_{0 < k \leq 1}, v \rangle,$$

where W and V are the same ones as in the standard conditional models. g_k is defined by a fuzzy measure m as

$$g_k(t, X) \stackrel{\text{def}}{=} \{Y \subseteq 2^W \mid m(Y, X) \geq k\}. \blacksquare$$

The model $\mathcal{M}_{\text{gCL}}^m$ is called *finite* because so is W . Further, in this paper, we call the model $\mathcal{M}_{\text{gCL}}^m$ *uniform* since functions $\{g_k\}$ in the model does not depend on any world in $\mathcal{M}_{\text{gCL}}^m$.

The truth conditions for $\Box \rightarrow_k$ and $\Diamond \rightarrow_k$ in a measure-based conditional model are given by

$$\begin{aligned} \mathcal{M}_{\text{gCL}}^m, t \models p \Box \rightarrow_k q &\text{ iff } \|q\|_{\mathcal{M}_{\text{gCL}}^m} \in g_k(t, \|p\|_{\mathcal{M}_{\text{gCL}}^m}), \\ \mathcal{M}_{\text{gCL}}^m, t \models p \Diamond \rightarrow_k q &\text{ iff } (\|q\|_{\mathcal{M}_{\text{gCL}}^m})^C \notin g_k(t, \|p\|_{\mathcal{M}_{\text{gCL}}^m}). \end{aligned}$$

The basic idea of these definitions is the same as in fuzzy-measure-based semantics for graded modal logic defined in [12–14].

When we take m as a conditional probability, the truth conditions of graded conditional becomes

$$\mathcal{M}_{\text{gCL}}^{\text{Pr}}, t \models p \Box \rightarrow_k q \text{ iff } \Pr(\|q\|_{\mathcal{M}_{\text{gCL}}^{\text{Pr}}} \mid \|p\|_{\mathcal{M}_{\text{gCL}}^{\text{Pr}}}) \geq k.$$

We have several soundness results based on probability-measure-based semantics (cf. [12–14]) shown in Table 9.1.

Table 9.1 Soundness results of graded conditionals by probability measures

$0 < k \leq \frac{1}{2}$	$\frac{1}{2} < k < 1$	$k = 1$	Rules and axiom schemata
○	○	○	RCEA. $\frac{p \leftrightarrow q}{(p \Box \rightarrow_k q) \leftrightarrow (q \Box \rightarrow_k q)}$
○	○	○	RCEC. $\frac{q \leftrightarrow q'}{(p \Box \rightarrow_k q) \leftrightarrow (p \Box \rightarrow_k q')}$
○	○	○	RCM. $\frac{q \leftrightarrow q'}{(p \Box \rightarrow_k q) \rightarrow (p \Box \rightarrow_k q')}$
		○	RCR. $\frac{(q \wedge q') \rightarrow r}{((p \Box \rightarrow_k q) \wedge (p \Box \rightarrow_k q')) \rightarrow (p \Box \rightarrow_k r)}$
○	○	○	RCN. $\frac{q}{p \Box \rightarrow_k q}$
		○	RCK. $\frac{(q_1 \wedge \dots \wedge q_n) \rightarrow q}{((p \Box \rightarrow_k q_1) \wedge \dots \wedge (p \Box \rightarrow_k q_n)) \rightarrow (p \Box \rightarrow_k q)}$
○	○	○	CM. $(p \Box \rightarrow_k (q \wedge r)) \rightarrow (p \Box \rightarrow_k q) \wedge (p \Box \rightarrow_k r)$
		○	CC. $(p \Box \rightarrow_k q) \wedge (p \Box \rightarrow_k r) \rightarrow (p \Box \rightarrow_k (q \wedge r))$
		○	CR. $(p \Box \rightarrow_k (q \wedge r)) \leftrightarrow (p \Box \rightarrow_k q) \wedge (p \Box \rightarrow_k r)$
○	○	○	CN. $p \Box \rightarrow_k \top$
○	○	○	CP. $\neg(p \Box \rightarrow_k \perp)$
		○	CK. $(p \Box \rightarrow_k (q \rightarrow r)) \rightarrow (p \Box \rightarrow_k q) \rightarrow (p \Box \rightarrow_k r)$
	○	○	CD. $\neg((p \Box \rightarrow_k q) \wedge (p \Box \rightarrow_k \neg q))$
○			CD_C. $(p \Box \rightarrow_k q) \vee (p \Box \rightarrow_k \neg q)$

9.3 Paraconsistency and Paracompleteness in Conditionals

As Chellas pointed out in his book [7] (p. 269), conditionals $p \Box \rightarrow q$ (and also $p \Diamond \rightarrow q$) is regarded as relative modal sentences like $[p]q$ (and also $\langle p \rangle q$). So we first see paraconsistency and paracompleteness in the usual modal setting for convenience.

9.3.1 Modal Logic Case

Let us define some standard language \mathcal{L} for modal logic with two modal operators \Box and \Diamond . In [18], we examined some relationship between modal logics and paraconsistency and paracompleteness. Let us assume a language \mathcal{L} of modal logic as usual. In terms of modal logic, paracompleteness and paraconsistency have a close relation to the following axiom schemata:

$$\begin{aligned} \mathbf{D}. & \quad \Box p \rightarrow \neg \Box \neg p, \\ \mathbf{D}_C. & \quad \neg \Box \neg p \rightarrow \Box p, \end{aligned}$$

because they have their equivalent expressions

$$\begin{aligned} & \neg(\Box p \wedge \Box \neg p), \\ & \Box p \vee \Box \neg p, \end{aligned}$$

respectively. That is, given a system of modal logic Σ , define the following set of sentences

$$T \stackrel{\text{def}}{=} \{p \in \mathcal{L} \mid \vdash_{\Sigma} \Box p\},$$

where $\vdash_{\Sigma} \Box p$ means $\Box p$ is a theorem of Σ . Then the above two schemata mean that, for any sentence p

$$\begin{aligned} & \text{not}(p \in T \text{ and } \neg p \in T) \\ & p \in T \text{ or } \neg p \in T \end{aligned}$$

respectively, and obviously the former describes the consistency of T and the latter the completeness of T . Thus

- T is inconsistent when Σ does not contain **D**.
- T is incomplete when Σ does not contain **D_C**.

A system Σ is regular when it contains the following rule and axiom schemata

$$\begin{aligned} & p \leftrightarrow q \Rightarrow \Box p \leftrightarrow \Box q \\ & (\Box p \wedge \Box q) \leftrightarrow \Box(p \wedge q) \end{aligned}$$

Note that any normal system is regular.

In [18] we pointed out the followings. If Σ is regular, then we have

$$(\Box p \wedge \Box \neg p) \leftrightarrow \Box \neg \top \quad (9.1)$$

where $\perp \leftrightarrow \neg \top$ and \perp is falsity constant, which means inconsistency itself. Thus we have triviality:

$$T = \mathcal{L}.$$

But if Σ is not regular, then we have no longer (9.1), thus, in general

$$T \neq \mathcal{L},$$

which means T is paraconsistent. That is, local inconsistency does not generate triviality as global inconsistency.

9.3.2 Conditional Logic Case

Next we apply the previous idea into conditional logics. In conditional logics, the corresponding axiom schemata

$$\begin{aligned} \mathbf{CD.} & \quad \neg((p \Box \rightarrow q) \wedge (p \Box \rightarrow \neg q)) \\ \mathbf{CD}_C. & \quad (p \Box \rightarrow q) \vee (p \Box \rightarrow \neg q) \end{aligned}$$

Given a system CL of conditional logic, define the following set of conditionals (rules):

$$R \stackrel{\text{def}}{=} \{p \Box \rightarrow q \in \mathcal{L}_{CD} \mid \vdash_{CL} p \Box \rightarrow q\}.$$

where \mathcal{L}_{CD} is a language for conditional logic and $\vdash_{CL} p \Box \rightarrow q$ means $p \Box \rightarrow q$ is a theorem of CL . Then the above two schemata mean that, for any sentence p

$$\begin{aligned} & \text{not}(p \Box \rightarrow q \in R \text{ and } p \Box \rightarrow \neg q \in R) \\ & p \Box \rightarrow q \in R \text{ or } p \Box \rightarrow \neg q \in R \end{aligned}$$

respectively, and obviously the former describes the consistency of R and the latter the completeness of R . Thus, for the set R of conditionals (rules)

- R is inconsistent when CL does not contain \mathbf{CD} .
- R is incomplete when CL does not contain \mathbf{CD}_C .

9.4 Paraconsistency and Paracompleteness in Association Rules

9.4.1 Association Rules

Let \mathcal{I} be a finite set of *items*. Any subset X in \mathcal{I} is called an *itemset* in \mathcal{I} . A database is comprised of *transactions*, which are actually obtained or observed itemsets. Formally, we give the following definition:

Definition 9.4 A database \mathcal{D} on \mathcal{I} is defined as

$$\langle T, V \rangle,$$

where

1. $T = \{1, 2, \dots, n\}$ (n is the size of the database),
2. $V : T \rightarrow 2^{\mathcal{I}}$. ■

Thus, for each transaction $i \in T$, V gives its corresponding set of items $V(i) \subseteq \mathcal{I}$.

For an itemset X , its *degree of support* $s(X)$ is defined by

$$s(X) \stackrel{\text{def}}{=} \frac{|\{t \in T \mid X \subseteq V(t)\}|}{|T|},$$

where $|\cdot|$ is a size of a finite set.

Definition 9.5 (Agrawal et al. [1]) Given a set of items \mathcal{I} and a database \mathcal{D} on \mathcal{I} , an *association rule* is an implication of the form

$$X \implies Y,$$

where X and Y are itemsets in \mathcal{I} with $X \cap Y = \emptyset$. ■

The following two indices were introduced in [1].

Definition 9.6 (Agrawal et al. [1])

1. An association rule $r = (X \implies Y)$ holds with *confidence* $c(r)$ ($0 \leq c(r) \leq 1$) in \mathcal{D} if and only if

$$c(r) = \frac{s(X \cup Y)}{s(X)}.$$

2. An association rule $r = (X \implies Y)$ has a *degree of support* $s(r)$ ($0 \leq s(r) \leq 1$) in \mathcal{D} if and only if

$$s(r) = s(X \cup Y). \quad \blacksquare$$

In this paper, we will deal with the former index.

Mining of association rules is actually performed by generating all rules that have certain minimum support (denoted *minsup*) and minimum confidence (denoted *minconf*) that a user specifies. See, e.g., [1–3] for details of such algorithms for finding association rules.

For example, consider the movie database in Table 9.2, where AH and HM means Ms. Audrey Hepburn and Mr. Henry Mancini, respectively. If you have watched several (famous) Ms. Hepburn’s movies, you might hear some wonderful music composed by Mr. Mancini. This can be represented by the association rule

$$r = \{AH\} \implies \{HM\}$$

with its confidence

$$c(r) = \frac{s(\{AH\} \cup \{HM\})}{s(\{AH\})} = 0.5$$

and its degree of support

$$s(r) = \frac{|\{T \mid T \subseteq \{AH\} \cup \{HM\}\}|}{|D|} = \frac{4}{100} = 0.04.$$

Table 9.2 Movie database

No.	Transaction (movie)	AH	HM
1	Secret people	1	
2	Monte Carlo baby	1	
3	Roman holiday	1	
4	My fair lady	1	
5	Breakfast at Tiffany’s	1	1
6	Charade	1	1
7	Two for the road	1	1
8	Wait until dark	1	1
9	Days of wine and rose		1
10	The great race		1
11	The pink panther		1
12	Sunflower		1
13	Some like it hot		
14	12 Angry men		
15	The apartment		
		
100	Les aventuriers		

9.4.2 Measure-Based Conditional Models for Databases

Let us regards a finite set \mathcal{I} of items as *atomic sentences*. Then, a *language* $\mathcal{L}_{\text{gCL}}(\mathcal{I})$ for graded conditional logic is formed from \mathcal{I} as the set of sentences closed under the usual propositional operators such as \top , \perp , \neg , \wedge , \vee , \rightarrow , and \leftrightarrow as well as $\Box \rightarrow_k$ and $\Diamond \rightarrow_k$ (*graded conditionals*) for $0 < k \leq 1$ in the usual way.

1. If $x \in \mathcal{I}$ then $x \in \mathcal{L}_{\text{gCL}}(\mathcal{I})$.
2. $\top, \perp \in \mathcal{L}_{\text{gCL}}(\mathcal{I})$.
3. If $p \in \mathcal{L}_{\text{gCL}}(\mathcal{I})$ then $\neg p \in \mathcal{L}_{\text{gCL}}(\mathcal{I})$.
4. If $p, q \in \mathcal{L}_{\text{gCL}}(\mathcal{I})$ then $p \wedge q, p \vee q, p \rightarrow q, p \leftrightarrow q \in \mathcal{L}_{\text{gCL}}(\mathcal{I})$,
5. If $[p, q \in \mathcal{L}_{\text{gCL}}(\mathcal{I})$ and $0 < k \leq 1]$ then $p \Box \rightarrow_k q, p \Diamond \rightarrow_k q \in \mathcal{L}_{\text{gCL}}(\mathcal{I})$.

A measure-based conditional model is defined as a family of minimal conditional model (cf. Chellas [7]):

Definition 9.7 Given a database $\mathcal{D} = \langle T, V \rangle$ on \mathcal{I} and a fuzzy measure m , a *measure-based conditional model* $\mathcal{M}_{\text{g}\mathcal{D}}^m$ for \mathcal{D} is a structure

$$\langle W, \{\mathbf{g}_k\}_{0 < k \leq 1}, v \rangle,$$

where (1) $W = T$, (2) for any world (transaction) t in W and any set of itemsets X in $2^{\mathcal{I}}$, \mathbf{g}_k is defined by a fuzzy measure m as

$$\mathbf{g}_k(t, X) \stackrel{\text{def}}{=} \{Y \subseteq 2^W \mid m(Y, X) \geq k\},$$

and (3) for any item x in \mathcal{I} , $v(x, t) = 1$ iff $x \in V(t)$. ■

The model $\mathcal{M}_{\text{g}\mathcal{D}}^m$ is called *finite* because so is W . Further, in this paper, we call the model $\mathcal{M}_{\text{g}\mathcal{D}}^m$ *uniform* since functions $\{\mathbf{g}_k\}$ in the model does not depend on any world in $\mathcal{M}_{\text{g}\mathcal{D}}^m$.

The truth condition for $\Box \rightarrow_k$ in a grade conditional model is given by

$$\mathcal{M}_{\text{g}\mathcal{D}}^m, t \models p \Box \rightarrow_k q \text{ iff } \|q\|_{\mathcal{M}_{\text{g}\mathcal{D}}^m} \in \mathbf{g}_k(t, \|p\|_{\mathcal{M}_{\text{g}\mathcal{D}}^m}),$$

where

$$\|p\|_{\mathcal{M}_{\text{g}\mathcal{D}}^m} \stackrel{\text{def}}{=} \{t \in W (= T) \mid \mathcal{M}_{\text{g}\mathcal{D}}^m, t \models p\}.$$

The basic idea of this definition is the same as in fuzzy-measure-based semantics for graded modal logic defined in [12–14].

9.4.3 Association Rules and Graded Conditionals

For example, the usual degree of confidence [1] is nothing but the well-known conditional probability, so we define function \mathbf{g}_k by conditional probability.

Definition 9.8 For a given database $\mathcal{D} = \langle T, V \rangle$ on \mathcal{I} and a conditional probability

$$\Pr(B|A) = \frac{|A \cap B|}{|A|},$$

its corresponding *probability-based graded conditional model* $\mathcal{M}_{\mathcal{D}}^{\text{Pr}}$ is defined as a structure

$$\langle W, \{g_k\}_{0 < k \leq 1}, v \rangle,$$

where

$$g_k(w, X) \stackrel{\text{def}}{=} \{Y \subseteq 2^W \mid \Pr(t(Y) \mid t(X)) \geq k\},$$

where

$$t(X) \stackrel{\text{def}}{=} \{w \in W \mid X \subseteq w\}. \blacksquare$$

The truth condition of graded conditional is given by

$$\mathcal{M}_{\mathcal{D}}^{\text{Pr}}, t \models p \Box \rightarrow_k q \text{ iff } \Pr(\|q\|_{\mathcal{M}_{\mathcal{D}}^{\text{Pr}}} \mid \|p\|_{\mathcal{M}_{\mathcal{D}}^{\text{Pr}}}) \geq k.$$

Then we can have the following theorem:

Theorem 9.1 Given a database \mathcal{D} on \mathcal{I} and its corresponding probability-based graded conditional model $\mathcal{M}_{\mathcal{D}}$, for an association rule $X \Longrightarrow Y$, we have

$$c(X \Longrightarrow Y) \geq k \text{ iff } \mathcal{M}_{\mathcal{D}}^{\text{Pr}} \models p_X \Box \rightarrow_k p_Y. \blacksquare$$

9.4.4 Paraconsistency and Paracompleteness in Association Rules

We formulated association rules as graded conditionals based on probability. Define the following set of rules with confidence k :

$$R_k \stackrel{\text{def}}{=} \{p \Box \rightarrow_k q \in \mathcal{L}_{gCD} \mid \vdash_{gCL} p \Box \rightarrow_k q\}.$$

A graded conditional $p \Box \rightarrow_k q$ is also regarded as a relative necessary sentences $[p]_k q$ and the properties of relative modal operator $[\cdot]_k$ are examined in Murai et al. [12, 13], [14] in the following correspondence:

Confidence k	Systems
$0 < k \leq \frac{1}{2}$	EMD_CNP
$\frac{1}{2} < k < 1$	$EMDNP$
$k = 1$	KD

The former two systems are not regular, so R_k may be paraconsistent. The last one is normal so regular.

For $0 < k \leq \frac{1}{2}$, R_k is complete but for some p and q , the both rules $p \Box \rightarrow_k q$ and $p \Box \rightarrow_k \neg q$ may be generated. This should be avoided.

For $\frac{1}{2} < k < 1$, R_k is consistent but may be paracomplete.

9.5 Dempster-Shafer-Theory-Based Confidence

9.5.1 D-S Theory and Confidence

The standard confidence [1] described in the previous section is based on the idea that co-occurrences of items in one transaction are evidence for association between the items. Since the definition of confidence is nothing but a conditional probability, even weights are a priori assigned to each transaction that contains the items in question at the same time. All of such transactions, however, do not necessarily give us such evidence because some co-occurrences might be contingent. Thus we need a framework that can differentiate proper evidence from contingent one and we introduce Dempster-Shafer theory of evidence [9, 19] to describe such a more flexible framework to compute confidence. There are a variety of ways of formalizing D-S theory and, in this paper, we adopt multivalued-mapping-based approach, which was originally used by Dempster [9]. In the approach, we need two frames, one of which has a probability defined, and a multivalued mapping between the two frames. Given a database $\mathcal{D} = \langle T, V \rangle$ on \mathcal{I} and an association rule $r = (X \implies Y)$ in \mathcal{D} , one of frames is the set T of transactions. Another one is defined by

$$R = \{r, \bar{r}\},$$

where \bar{r} denotes the negation of r . The remaining tasks are (1) to define a probability distribution Pr on T : $\text{Pr} : T \rightarrow [0, 1]$, and (2) to define a multivalued mapping $\Gamma : T \rightarrow 2^R$. Given Pr and Γ , we can define the well-known two kinds of functions in Dempster-Shafer theory: for $X \subseteq 2^R$,

$$\begin{aligned} \text{Bel}(X) &\stackrel{\text{def}}{=} \text{Pr}(\{t \in T \mid \Gamma(t) \subseteq X\}), \\ \text{Pl}(X) &\stackrel{\text{def}}{=} \text{Pr}(\{t \in T \mid \Gamma(t) \cap X \neq \emptyset\}) \end{aligned}$$

which are called *belief* and *plausibility* functions, respectively. Now we have the following double-indexed confidence:

$$c(r) = \langle \text{Bel}(r), \text{Pl}(r) \rangle.$$

9.5.2 Multi-graded Conditional Models for Databases

Given a finite set \mathcal{I} of items as *atomic sentences*, a language $\mathcal{L}_{\text{mgCL}}(\mathcal{I})$ for graded conditional logic is formed from \mathcal{I} as the set of sentences closed under the usual propositional operators as well as $\Box \rightarrow_k$ and $\Diamond \rightarrow_k$ (*graded conditionals*) for $0 < k \leq 1$ in the usual way. Note that, in particular,

$$(p, q \in \mathcal{L}_{\text{mgCL}}(\mathcal{I}) \text{ and } 0 < k \leq 1) \Rightarrow p\Box \rightarrow_k q, p\Diamond \rightarrow_k q \in \mathcal{L}_{\text{mgCL}}(\mathcal{I}).$$

Definition 9.9 Given a database \mathcal{D} on \mathcal{I} , a *multi-graded conditional model* $\mathcal{M}_{\text{mg}\mathcal{D}}$ for \mathcal{D} is a structure

$$\langle W, \{\{\underline{g}_k, \bar{g}_k\}\}_{0 < k \leq 1}, v \rangle,$$

where (1) $W = T$, (2) for any world (transaction) t in W and any set of itemsets \mathcal{X} in $2^{\mathcal{I}}$, \underline{g}_k is defined by belief and plausibility functions:

$$\begin{aligned} \underline{g}_k(t, \mathcal{X}) &\stackrel{\text{def}}{=} \{\mathcal{Y} \subseteq 2^W \mid \text{Bel}(\mathcal{Y}, \mathcal{X}) \geq k\}, \\ \bar{g}_k(t, \mathcal{X}) &\stackrel{\text{def}}{=} \{\mathcal{Y} \subseteq 2^W \mid \text{Pl}(\mathcal{Y}, \mathcal{X}) \geq k\}, \end{aligned}$$

and (3) for any item x in \mathcal{I} , $v(x, t) = 1$ iff $x \in V(t)$ ■

The truth conditions for $\Box \rightarrow_k$ and $\Diamond \rightarrow_k$ are given by

$$\begin{aligned} \mathcal{M}_{\text{mg}\mathcal{D}}, w \models p\Box \rightarrow_k q &\text{ iff } \|q\|^{\mathcal{M}_{\text{mg}\mathcal{D}}} \in \underline{g}_k(t, \|p\|^{\mathcal{M}_{\text{mg}\mathcal{D}}}) \\ \mathcal{M}_{\text{mg}\mathcal{D}}, w \models p\Diamond \rightarrow_k q &\text{ iff } \|q\|^{\mathcal{M}_{\text{mg}\mathcal{D}}} \in \bar{g}_k(t, \|p\|^{\mathcal{M}_{\text{mg}\mathcal{D}}}), \end{aligned}$$

respectively. Its basic idea is also the same as in fuzzy-measure-based semantics for graded modal logic defined in [12–14]. Several soundness results based on belief- and plausibility-function-based semantics (cf. [12–14]) are shown in Table 9.3.

9.5.3 Two Typical Cases

First we define a probability distribution on T by

$$\text{Pr}(t) \stackrel{\text{def}}{=} \begin{cases} \frac{1}{a}, & \text{if } t \in \|p_X\|^{\mathcal{M}_{\text{mg}\mathcal{D}}}, \\ 0, & \text{otherwise,} \end{cases}$$

where $a = \|\|p_X\|^{\mathcal{M}_{\text{mg}\mathcal{D}}}\|$. This means that each world (transaction) t in $\|p_X\|^{\mathcal{M}_{\text{mg}\mathcal{D}}}$ is given an even mass (weight) $\frac{1}{a}$. To generalize the distribution is of course another interesting task.

Table 9.3 Soundness results of graded conditionals by belief and plausibility functions

Belief function			Rules and axiom schemata	Plausibility function		
$0 < k \leq \frac{1}{2}$	$\frac{1}{2} < k < 1$	$k = 1$		$0 < k \leq \frac{1}{2}$	$\frac{1}{2} < k < 1$	$k = 1$
○	○	○	RCEA	○	○	○
○	○	○	RCEC	○	○	○
○	○	○	RCM	○	○	○
○	○	○	RCR			○
○	○	○	RCN	○	○	○
		○	RCK			
○	○	○	CM	○	○	○
○	○	○	CC			○
○	○	○	CR			○
○	○	○	CN	○	○	○
○	○	○	CP	○	○	○
		○	CK			
	○	○	CD			
			CD_C	○		

Next we shall see two typical cases of definition of Γ . First we describe strongest cases. When we define a mapping Γ by

$$\Gamma(t) \stackrel{\text{def}}{=} \begin{cases} \{r\}, & \text{if } t \in \|p_X\|^{\mathcal{M}_{\text{mg}\mathcal{D}}}, \\ \{\bar{r}\}, & \text{otherwise.} \end{cases}$$

This means that the transactions in $\|p_X \wedge p_Y\|^{\mathcal{M}_{\text{mg}\mathcal{D}}}$ contribute as evidence to r , while the transactions in $\|p_X \wedge \neg p_Y\|^{\mathcal{M}_{\text{mg}\mathcal{D}}}$ contribute as evidence to \bar{r} . This is the strongest interpretation of co-occurrences. Then, we can compute $\text{Bel}(r) = \frac{1}{a} \times b$ and $\text{Pl}(r) = \frac{1}{a} \times b$, where $b = \|\|p_X \wedge p_Y\|^{\mathcal{M}_{\text{mg}\mathcal{D}}}\|$. Thus the induced belief and plausibility functions collapse to the same probability measure Pr : $\text{Bel}(r) = \text{Pl}(r) = \text{Pr}(r) = \frac{b}{a}$, and thus

$$c(r) = \left\langle \frac{b}{a}, \frac{b}{a} \right\rangle.$$

Hence this case represents the usual confidence. According to this idea, in our movie database, we can define Pr and Γ in the way in Fig. 9.1. That is, any movie in $\|\text{AH} \wedge \text{HM}\|^{\mathcal{M}_{\text{mg}\mathcal{D}}}$ contributes as evidence to that the rule holds (r), while all movie in $\|\text{AH} \wedge \neg \text{HM}\|^{\mathcal{M}_{\text{mg}\mathcal{D}}}$ contributes as evidence to that the rule does not hold (\bar{r}). Thus we have

$$c(\{\text{AH}\} \implies \{\text{HM}\}) = \langle 0.5, 0.5 \rangle.$$

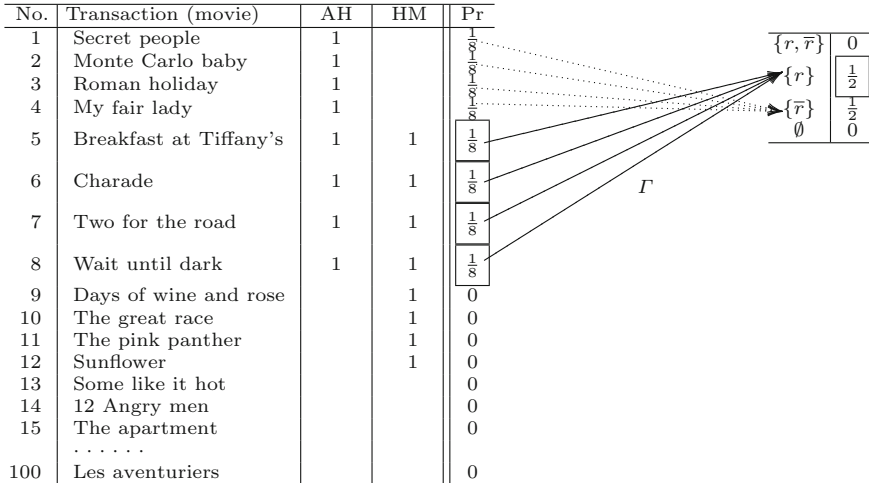


Fig. 9.1 An example of the strongest cases

Next we describe weakest cases. In general, co-occurrences do not necessarily mean actual association. The weakest interpretation of co-occurrences is to consider transactions totally unknown as described as follows: When we define a mapping Γ by

$$\Gamma(t) \stackrel{\text{def}}{=} \begin{cases} \{r, \bar{r}\}, & \text{if } t \in \|p_X\|^{\mathcal{M}_{\text{mgD}}}, \\ \{\bar{r}\}, & \text{otherwise.} \end{cases}$$

This means that the transactions in $\|p_X \wedge p_Y\|^{\mathcal{M}_{\text{mgD}}}$ contribute as evidence to $R = \{r, \bar{r}\}$, while the transactions in $\|p_X \wedge \neg p_Y\|^{\mathcal{M}_{\text{mgD}}}$ contribute as evidence to \bar{r} . Then, we can compute $\text{Bel}(r) = 0$ and $\text{Pl}(r) = \frac{1}{a} \times b$, and thus

$$c(r) = \langle 0, \frac{b}{a} \rangle.$$

According to this idea, in our movie database, we can define Pr and Γ in the way in Fig. 9.2. That is, all movie in $\|AH \wedge \neg HM\|^{\mathcal{M}_{\text{mgD}}}$ contributes as evidence to that the rule does not hold (\bar{r}), while we cannot expect whether each movie in $\|AH \wedge HM\|^{\mathcal{M}_{\text{mgD}}}$ contributes or not as evidence to that the rule holds (r). Thus we have

$$c(\{AH\} \implies \{HM\}) = \langle 0, 0.5 \rangle.$$

In the case, the induced belief and plausibility functions, denoted respectively $\text{Bel}_{bpa'}$ and $\text{Pl}_{bpa'}$, become *necessity* and *possibility* measures in the sense of Dubois and Prade [10]. We have several soundness results based on necessity- and possibility-measure-based semantics (cf. [12–14]) shown in Table 9.4.

No.	Transaction (movie)	AH	HM	Pr
1	Secret people	1		
2	Monte Carlo baby	1		
3	Roman holiday	1		
4	My fair lady	1		
5	Breakfast at Tiffany's	1	1	$\frac{1}{2}$
6	Charade	1	1	$\frac{1}{2}$
7	Two for the road	1	1	$\frac{1}{2}$
8	Wait until dark	1	1	$\frac{1}{2}$
9	Days of wine and rose		1	0
10	The great race		1	0
11	The pink panther		1	0
12	Sunflower		1	0
13	Some like it hot			0
14	12 Angrymen			0
15	The apartment			0
			
100	Les aventuriers			0

Fig. 9.2 An example of the weakest cases

Table 9.4 Soundness results of graded conditionals by necessity and possibility measures

Necessity measure $0 < k \leq 1$	Rules and axiom schemata	Possibility measure $0 < k \leq 1$
○	RCEA	○
○	RCEC	○
○	RCM	○
○	RCR	
○	RCN	○
○	RCK	
○	CM	○
○	CC	
	CF	○
○	CR	○
○	CN	○
○	CP	○
○	CK	
○	CD	
	CD_c	○

9.5.4 General Cases

In the previous two typical cases, one of which coincides to the usual confidence, any transaction in $\|AH \wedge HM\|^{\mathcal{M}_{mg\mathcal{D}}}$ (or in $\|AH \wedge \neg HM\|^{\mathcal{M}_{mg\mathcal{D}}}$) has the same weight as evidence. It would be, however, possible that some of $\|AH \wedge HM\|^{\mathcal{M}_{mg\mathcal{D}}}$

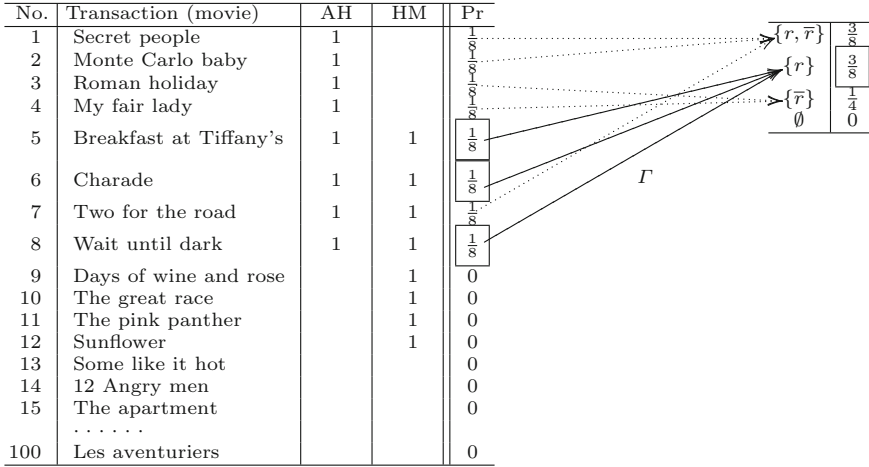


Fig. 9.3 An example of general cases

(or $\|AH \wedge \neg HM\|^{\mathcal{M}_{mgD}}$) does work as positive evidence to r (or \bar{r}) but other part does not.

Thus we have a tool that allows us to introduce various kinds of ‘a posteriori’ pragmatic knowledge into the logical setting of association rules. As an example, we assume that (1) the music of the first and second movies was not composed by Mancini, but the fact does not affect the validity of \bar{r} because they are not very important ones, and (2) the music of the seventh movie was composed by Mancini, but the fact does not affect the validity of r . Then we can define Γ in the way in Fig. 9.3. Thus we have

$$c(\{AH\} \implies \{HM\}) = \langle 0.375, 0.75 \rangle.$$

In general, users have such kind of knowledge ‘a posteriori.’ Thus the D-S based approach allows us to introduce various kinds of ‘a posteriori’ pragmatic knowledge into association rules.

9.6 Concluding Remarks

In this paper, we examined paraconsistency and paracompleteness that appear in association rules in a framework of probability-based models for conditional logics. For lower values of confidence (less than or equal to $\frac{1}{2}$), both $p \square \rightarrow_k q$ and $p \square \rightarrow_k \neg q$ may be generated so we must be careful to use such lower confidence.

Further we extended the above discussion into the case of Dempster-Shafer theory of evidence to double-indexed confidences. Users have, in general, such kind

of knowledge ‘a posteriori’ describe in the previous section. Thus the D-S based approach allows a sophisticated way of calculating confidence by introducing various kinds of ‘a posteriori’ pragmatic knowledge into association rules.

Acknowledgments We are grateful to a referee for useful comments.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD Conference on Management of Data, pp. 207–216 (1993)
2. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.I.: Fast discovery of association rules. In: Fayyad, U.M., Platetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds.) *Advances in Knowledge Discovery and Data Mining*, pp. 307–328. AAAI Press/The MIT Press (1996)
3. Aggarwal, C.C., Philip, S.Y.: Online generation of association rules. In: Proceedings of the International Conference on Data Engineering, pp. 402–411 (1998)
4. Akama, S., Abe, J.M.: Many-valued and annotated modal logics. In: Proceedings of 28th ISMVL, pp. 114–119 (1998)
5. Akama, S., Abe, J.M.: Fuzzy annotated logics. In: Proceedings of IPMU 2000, pp. 504–509 (2000)
6. Akama, S., Abe, J.M.: Paraconsistent logics viewed as a foundation of data warehouses. *Advances in Logic, Artificial Intelligence and Robotics*, pp. 96–103. IOS Press (2002)
7. Chellas, B.F.: *Modal Logic: An Introduction*. Cambridge University Press, Cambridge (1980)
8. da Costa, N.C.A., Abe, J.M., Subrahmanian, V.S.: Remarks on annotated logic. *Zeitschr. f. Math. Logik und Grundlagen d. Math.* **37**, 561–570 (1991)
9. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.* **38**, 325–339 (1967)
10. Dubois, D., Prade, H.: *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Springer (1988)
11. Lewis, D.: *Counterfactuals*. Blackwell, Oxford (1973)
12. Murai, T., Miyakoshi, M., Shimbo, M.: Measure-based semantics for modal logic. In: Lowen, R., Roubens, M. (eds.) *Fuzzy Logic: State of the Art*, pp. 395–405. Kluwer, Dordrecht (1993)
13. Murai, T., Miyakoshi, M., Shimbo, M.: Soundness and completeness theorems between the Dempster-Shafer theory and logic of belief. In: Proceedings of the 3rd FUZZ-IEEE (WCCI), pp. 855–858 (1994)
14. Murai, T., Miyakoshi, M., Shimbo, M.: A logical foundation of graded modal operators defined by fuzzy measures. In: Proceedings of the 4th FUZZ-IEEE/2nd IFES, pp. 151–156 (1995)
15. Murai, T., Sato, Y.: Association rules from a point of view of modal logic and rough sets. In: Proceedings of the 4th AFSS, pp. 427–432 (2000)
16. Murai, T., Nakata, M., Sato, Y.: A note on conditional logic and association rules. In: Terano, T., et al. (eds.) *New Frontiers in Artificial Intelligence, LNAI*, vol. 2253, pp. 390–394. Springer (2001)
17. Murai, T., Nakata, M., Sato, Y.: Association rules as relative modal sentences based on conditional probability. *Commun. Inst. Inf. Comput. Mach.* **5**(2), 73–76 (2002)
18. Murai, T., Sato, Y., Kudo, Y.: Paraconsistency and neighborhood models in modal logic. In: Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics, vol. XII, pp. 220–223 (2003)
19. Shafer, G.: *A Mathematical Theory of Evidence*. Princeton University Press (1976)