Taxonomic and Uncertain Integrity Constraints in Object-Oriented Databases – the TOP Approach

Thomas Lukasiewicz[1] Werner Kießling[1] Gerhard Köstler[1] Ulrich Güntzer[2]

[1] Lehrstuhl für Informatik II, Universität Augsburg, Universitätsstr. 14, 86135 Augsburg, Germany, {lukasiewicz | kiessling | koestler}@uni-augsburg.de

[2] Wilhelm-Schickard-Institut, Universität Tübingen, Sand 13, 72076 Tübingen, Germany, guentzer@informatik.uni-tuebingen.de

Abstract

We present a coherent modeling and reasoning methodology to extend object-oriented databases towards taxonomic and uncertain integrity constraints. Our so-called TOP database model enriches current ISA-hierarchies by more general t-classes to improve conceptual modeling. The t-classes themselves are then integrated with probabilistic constraints to express uncertainty. We give an efficient algorithm for checking the modeling-consistency of a probabilistic knowledgebase. As a typical application domain for TOP, we exemplify how various aspects of a portfolio management system can be modeled. We also demonstrate that recent probabilistic inference methods, relying on a careful interaction between taxonomic and uncertain knowledge, can be applied in this context.

1 Introduction

The evolution of database technology from the relational model into object-oriented databases (OODBs) and deductive databases has been pushed forward to a state where stable and usable systems are becoming widely available; also an integration of both paradigms into so-called DOOD-systems is taking place. Such systems can deal with various kinds of objects and query mechanisms, however, except for attempts to deal with null-values, only the specification and processing of certain (true/false) information is supported so far. But uncertainty pervades the real world and it seems mandatory for future advanced data models to capture it explicitly and appropriately. Being a topic of interest in AI for quite some period of time, it is picked up by database researchers recently. There is e.g. work in the relational context by Barbará et al. [BGMP92] and for deductive databases by Ng and Subramanian [NS92], Lakshmanan and Sadri [LS94]. Our own previous work comprises a major project with the so-called DUCK-system for reasoning under a conditional probability model (see e.g. [GKT91], [TKG95]).

Before extending current OODBs towards uncertainty, the following aspects must be considered: Since uncertain knowledge comes in a variety of flavors in the real world, which of its facets should we provide? (See [Pea88], [Som90] for a discussion.) Like other database researchers we decided on the probabilistic model of uncertainty, which canonically extends the taxonomic interpretation of subtyping in OODBs as e.g. proposed by the ODMG-93 standard [Cat94]. The ODMG-93 standard optionally supports ISA-hierarchies, i.e. subtyping can be combined with an extensional interpretation of class-hierarchies yielding a set-inclusion order on the sets of objects assigned to all classes. This allows limited forms of classification by subclass-relationships, but not reaching the expressiveness of taxonomic classification found in terminological systems (see e.g. [Bra91]). On the other hand it is natural to understand taxonomic integrity constraints as special case of probabilistic knowledge. Consider the sentences "all dogs are domestic animals" and "at least 70% of all domestic animals are dogs" as examples of taxonomic and probabilistic knowledge, respectively.

Guided by these considerations, we propose an extension of OODBs towards the so-called *TOP database model*:

TOP = Taxonomy + Object-Orientation + Probability

The TOP-model shall provide a coherent knowledge representation schema which is compatible with the taxonomic view of OODB technology. Preliminary ideas of the TOP-model have been presented in [KLKG94]. In this paper we elaborate in more detail the integration of TOP-modeling aspects with recent theoretical results from the field of probabilistic deduction ([Luk95]).

The rest of this paper is organized as follows: Section 2 and 3 are concerned with the modeling aspects of TOP, introducing the notion of t-classes for formulating taxonomic and uncertain integrity constraints. Section 4 is dealing with a portfolio management application of the TOP database model. Section 5 applies recent results from the complex area of probabilistic deduction, highlighting the interplay with taxonomic deduction. Section 6 compares our results to related work and finally Sec. 7 gives a summary and an outlook on ongoing work.

2 Taxonomic integrity constraints

Taxonomic knowledge and reasoning is a widely explored field. One of its uses is in terminological reasoning to answer typical questions like "is a class of objects subset of or equal to some other classes of objects". ISA-hierarchies in OODBs express similar constraints on the extents of classes. However, except for the acyclic ISA-graphs supported, more

CIKM '95, Baltimore MD USA
© 1995 ACM 0-89791-812-6/95/11..$3.50

general features to express taxonomic integrity constraints are missing so far.

2.1 Syntax and semantics of taxonomic constraints

We start out with the definition of an expressive language for taxonomic classes, called t-classes.

Definition 2.1

a) We consider an alphabet $\mathcal{A} := \{\boldsymbol{\emptyset}, \boldsymbol{\mathcal{O}}, B_1, \ldots, B_k\}$ of constants. $\boldsymbol{\emptyset}$ is called *empty t-class term*, $\boldsymbol{\mathcal{O}}$ is called *universal t-class term*. $\mathcal{B} := \{B_1, \ldots, B_k\}$ denotes the set of *basic t-class terms*.

b) The set $\mathcal{C}$ of all *conjunctive t-class terms* is the minimal set with $\mathcal{A} \subseteq \mathcal{C}$ and $C, D \in \mathcal{C} \Longrightarrow (CD) \in \mathcal{C}$.

c) The set $\mathcal{G}$ of all *general t-class terms* is the minimal set with $\mathcal{A} \subseteq \mathcal{G}$ and $C, D \in \mathcal{G} \Longrightarrow (CD), (C \cup D), (\overline{C}) \in \mathcal{G}$.

□

For convenience we apply the usual preference rules to omit unnecessary parentheses.

Definition 2.2 Let $\mathcal{D}$ be the (not necessarily finite) set of admissible objects of an OODB schema $\mathcal{S}$. An *interpretation* $\mathcal{J} := (\mathcal{O}, J)$ of the set of general t-class terms $\mathcal{G}$ consists of:

a) A finite set of objects $\mathcal{O} \subseteq \mathcal{D}$ of an OODB over $\mathcal{S}$.

b) A mapping $J : \mathcal{B} \longrightarrow 2^{\mathcal{O}}$.

A basic t-class term B_i is interpreted as a class extension $J(B_i) = \{o_1, \ldots, o_l\}$ for objects $o_i \in \mathcal{O}$ and $l \geq 0$.

J is extended to $\mathcal{G}$ by $J(\boldsymbol{\mathcal{O}}) := \mathcal{O}$, $J(\boldsymbol{\emptyset}) := \emptyset$, $J(CD) := J(C) \cap J(D)$, $J(C \cup D) := J(C) \cup J(D)$ and $J(\overline{C}) := \mathcal{O} \backslash J(C)$.

In the sequel, we identify $\boldsymbol{\mathcal{O}}$ with $\mathcal{O}$, $\boldsymbol{\emptyset}$ with $\emptyset$ and B_i with $J(B_i)$. □

Classification among conjunctive t-classes can now be achieved by formulating *taxonomic constraints* in the following language.

Definition 2.3 Let $D_1, \ldots, D_m, D, C \in \mathcal{C}$.

a) $C \subseteq D$ is called a *subclass* constraint,

b) $C = D$ is called an *equality* constraint,

c) $D_1 \parallel \ldots \parallel D_m$ is called a *disjointness* constraint,

d) $D_1 \parallel \ldots \parallel D_m = D$ is called a *partition* constraint.

The set of all taxonomic constraints is denoted $\mathcal{TC}$. □

Definition 2.4 An interpretation $\mathcal{J} = (\mathcal{O}, J)$ of $\mathcal{G}$ is extended to an interpretation of $\mathcal{TC}$ into $\{true, false\}$ as follows:

a) $\mathcal{J} \models C \subseteq D$, iff $J(C) \subseteq J(D)$,

b) $\mathcal{J} \models C = D$, iff $\mathcal{J} \models C \subseteq D$ and $\mathcal{J} \models D \subseteq C$,

c) $\mathcal{J} \models D_1 \parallel \ldots \parallel D_m$, iff
$\mathcal{J} \models D_i D_j = \emptyset$ for all $i, j \in [1:m]$ with $i < j$,

d) $\mathcal{J} \models D_1 \parallel \ldots \parallel D_m = D$, iff
$\mathcal{J} \models D_1 \parallel \ldots \parallel D_m$ and $\mathcal{J} \models D_1 \cup \ldots \cup D_m = D$. □

The notions of models, satisfiability and logical consequence are defined as usual.

Definition 2.5

a) A *taxonomic knowledge-base* consists of a set of taxonomic constraints.

b) Let V be an infinite set of variables, let $F \in \mathcal{TC}$ and $A, B \in \mathcal{C} \cup V$.

The expressions $?F$, $?A \subseteq B$ and $?AB = \emptyset$ are called *taxonomic queries*. □

Since the interpretation $(\emptyset, \{(B_i, \emptyset) \mid i \in [1:k]\})$ is always a model, every taxonomic knowledge-base is satisfiable. But in the context of OODB modeling it is desirable to assure that the extension of every basic t-class can be different from $\emptyset$ and $\mathcal{O}$.

Definition 2.6 A taxonomic knowledge-base $\mathcal{T}$ is called *modeling-consistent*, iff there exists a model $(\mathcal{O}, J)$ of $\mathcal{T}$ with $J(B_i) \neq \emptyset$ and $J(B_i) \neq \mathcal{O}$ for all $B_i \in \mathcal{B}$. □

2.2 Extending OODBs by taxonomic constraints

From the various OODB-features, our interest here is focused on class hierarchies under the set-inclusion semantics. Other OODB-features such as aspects of attribute and method inheritance are not impacted by our subsequent considerations. Under our interpretation the relationship B_i *isa* B_j for two classes B_i and B_j of an OODB schema is equivalent to a subclass constraint $B_i \subseteq B_j$ with B_i and B_j as basic t-class terms. To make full taxonomic reasoning available as an extension of existing OODB technology, we propose the following two-step procedure:

Step 1: ISA-hierarchies of a conventional OODB schema are translated into $\mathcal{TC}$ according to Fig. 1. The diagrams are given in EER-notation, neglecting attributes and methods.

Step 2 (optional): Full $\mathcal{TC}$ is made available as a means to specify *additional* taxonomic constraints. This has no impact on the inheritance schema fixed before, it solely affects class extensions.

Example 2.7
Step 1: Let's consider an airline application and assume that the persons relevant for an air carrier are **passengers** and **employees**. These groups may not be disjoint, i.e., **employees** can be **passengers** as well. **Employees** are distinguished in pilots, **ground_staff**, **flight_attendants** and the remaining personnel. Subgroups of the **ground_staff** are **doctors** and **nurses**. They are not necessarily disjoint, since there might be doctors who are nurses as well.

After designing relevant attributes and methods, assume that the inheritance hierarchy is fixed (see Fig. 2, upper part).

Step 2: Additionally it is required that the group of **pilots** should be disjoint to all other groups of personnel. This is modeled by extra taxonomic constraints (see Fig. 2, lower part). □

It is crucial to observe that these two disjointness constraints could have been implemented by changing the original ISA-hierarchy radically by introducing an "artificial" superclass for **flight_attendants** and **ground_staff**. Obviously this procedure becomes very clumsy for larger numbers of subclasses and would render the conceptual model hard to

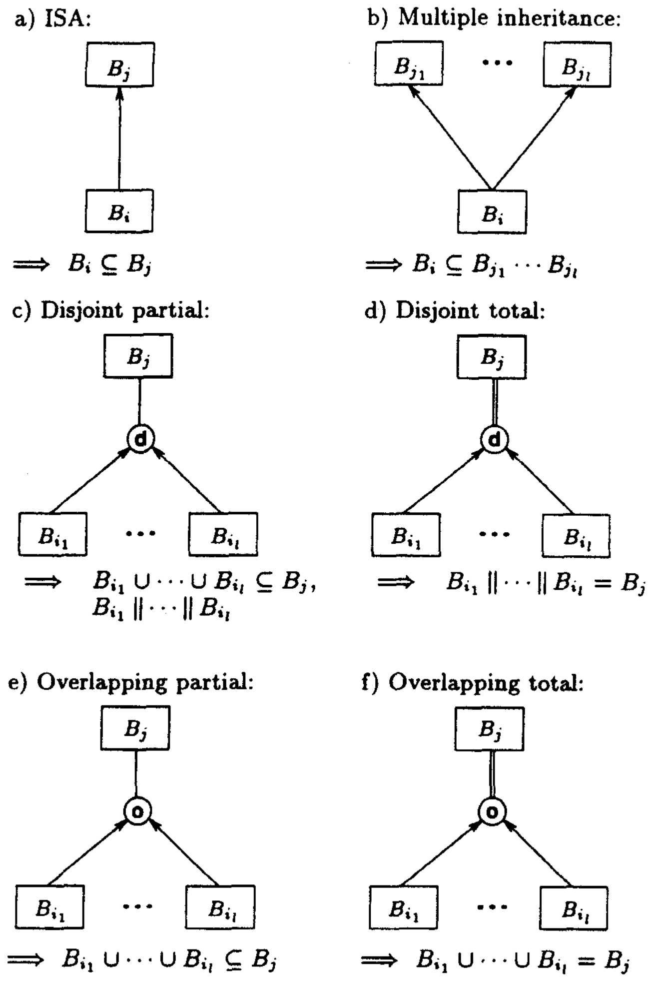

Figure 1: Some EER-constructs translated into $\mathcal{TC}$-constraints

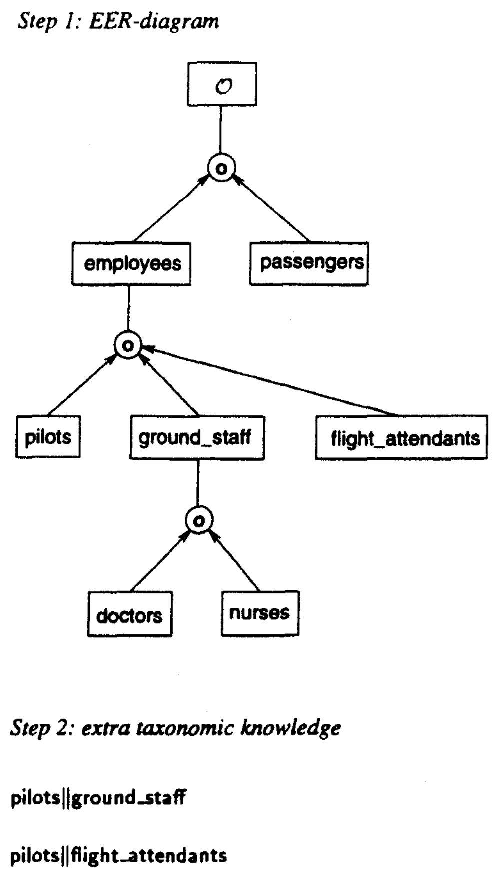

Figure 2: TOP-modeling of an airline application

understand. Moreover, the introduction of those superclasses does not possess any advantages w.r.t. software reuse, since by assumption in Step 1 all cases of useful inheritance had been fixed before. The TOP approach avoids such problems by decoupling inheritance and extra extensional taxonomic constraints carefully.

The conceptual modeling process, i.e., the design of the ISA-hierarchy and the formulation of additional taxonomic constraints, should be supported interactively by the system. To this end we can use taxonomic deduction to check the modeling-consistency (i.e., checking whether a class is forced to be empty) and to allow queries about the consequences of postulated taxonomic constraints. For instance, the user may wonder whether in Ex. 2.7 the groups of pilots and nurses are disjoint and pose the following taxonomic query:

$$?\text{nurses pilots} = \emptyset \ ,$$

yielding the answer *Yes*.
Another taxonomic query might be

$$?\text{nurses flight\_attendants} = \emptyset \ ,$$

yielding the answer *No*.

3 Uncertain integrity constraints

As announced, the TOP database model combines object-orientation and taxonomies with probability for modeling uncertainty. To achieve this we assume that we already have defined a taxonomic knowledge-base as described by the two-step procedure before. In a third step we now add uncertainty capabilities by a probabilistic model.

3.1 Syntax and semantics of probabilistic constraints

Probabilistic reasoning is a tremendously complex task with many intractability results known. Therefore many researchers a priori restrict their attention to tractable subclasses. In our previous work with the DUCK-approach ([GKT91], [TKG95]), we have tackled a very large class in which we subsequently identified important subclasses with sound, complete and efficient inference procedures. These subclasses are correlation programs ([TGK95]) and Bayesian networks with probabilistic intervals ([Thö94]). In the sequel we will focus on uncertain and correlation rules as *probabilistic constraints* in combination with taxonomic constraints.

Definition 3.1 Let $A, B \in \mathcal{C}$, $x_1, x_2, y_1, y_2 \in [0,1] \cap \mathbb{Q}$, $x_1 \leq x_2$ and $y_1 \leq y_2$.

a) $A \xrightarrow{x_1,x_2} B$ is called an *uncertain rule*,

b) $A \underset{y_1,y_2}{\overset{x_1,x_2}{\longleftrightarrow}} B$ is called a *correlation rule*, We abbreviate $A \xrightarrow{x,x} B$ by $A \xrightarrow{x} B$. The set of all probabilistic constraints is denoted $\mathcal{PC}$. □

Definition 3.2 An interpretation $\mathcal{J} = (\mathcal{O}, J, \mathrm{P})$ of $\mathcal{PC}$ consists of an interpretation $(\mathcal{O}, J)$ of $\mathcal{G}$ and a probability measure $\mathrm{P} : J(\mathcal{G}) \longrightarrow [0,1]$ for the measure space $(\mathcal{O}, J(\mathcal{G}))$.

a) $\mathcal{J} \models A \xrightarrow{x_1,x_2} B$, iff
$\mathrm{P}(J(A)) = 0$ or $x_1 \leq \mathrm{P}(J(B)|J(A)) \leq x_2$,[1]

b) $\mathcal{J} \models A \underset{y_1,y_2}{\overset{x_1,x_2}{\longleftrightarrow}} B$, iff

$$\mathcal{J} \models A \xrightarrow{x_1,x_2} B \text{ and } \mathcal{J} \models B \xrightarrow{y_1,y_2} A.$$ □

The notions of models, satisfiability and logical consequence are defined in the usual way.

Definition 3.3

a) A taxonomic knowledge-base, extended by a non-empty set of probabilistic constraints, is called a *probabilistic knowledge-base.*

b) Let V be an infinite set of variables. For $A, B \in \mathcal{C} \cup V$, $x_1, x_2 \in V$, the expression $?A \xrightarrow{x_1,x_2} B$ is called a *probabilistic query.*[2] □

Probabilistic knowledge-bases express user-defined (application-dependent) uncertain constraints for the frame of the "real world" to be modeled in an OODB. Note that in Def. 3.2 we do not commit ourselves to a specific interpretation of probability. The relative cardinality is just one possible interpretation. Its usage is illustrated by the following example.

Example 3.4 (Ex. 2.7 continued)
Step 3: In the EER-diagram of Fig. 2 **employees** and **passengers** are not modeled as disjoint classes, i.e., **passengers** may be **employees**. However, fictitiously assume because of economical reasons that their number should be restricted to 10 per cent. By company policy the number of **ground_staff**, **flight_attendants** and **pilots** is restricted to maximally 30, 20 and 5 per cent of the number of **employees**, resp., and the number of **doctors** and **nurses** is restricted to maximally 2 and 5 per cent of the number of **ground_staff**. Additionally assume that by legal regulations there must be at least 1 percent of **doctors** and 1 percent of **nurses** working in the **ground_staff** of an airline.

We model these restrictions by the following probabilistic constraints:

$$\text{passengers} \xrightarrow{0,0.1} \text{employees},$$
$$\text{employees} \xrightarrow{0,0.05} \text{pilots},$$
$$\text{employees} \xrightarrow{0,0.3} \text{ground\_staff},$$
$$\text{employees} \xrightarrow{0,0.2} \text{flight\_attendants},$$
$$\text{ground\_staff} \xrightarrow{0.01,0.02} \text{doctors},$$
$$\text{ground\_staff} \xrightarrow{0.01,0.05} \text{nurses}.$$ □

[1] $\mathrm{P}(J(B)|J(A))$ is the conditional probability of $J(B)$ under $J(A)$. The definition of uncertain rules incorporates the principle of "ex falso quodlibet", e.g. the uncertain rule flying elephants $\xrightarrow{0.7}$ helicopter is always true, since there are no flying elephants.

[2] Queries for correlation rules are treated similarly.

3.2 Integrating taxonomic and probabilistic constraints

In Sec. 2 we have introduced taxonomic knowledge-bases as a natural extension of ISA-hierarchies in OODBs. Probabilistic knowledge-bases enable us to represent *uncertain constraints* over object-oriented databases. However, we must be careful to smoothly integrate all pieces of knowledge to come up with a coherent knowledge representation schema. This is achieved by the following observation. For practical purpose it is sufficient and desirable to restrict the interpretation $(\mathcal{O}, J, \mathrm{P})$ of $\mathcal{PC}$ as follows:

Assumption 3.5 *We just consider interpretations $(\mathcal{O}, J, \mathrm{P})$ of $\mathcal{PC}$ with $\mathrm{P}(J(A)) = 0 \Longrightarrow J(A) = \emptyset$ for all $A \in \mathcal{C}$.*

This restriction naturally holds for a probabilistic interpretation by relative cardinalities. It entails the following correspondence between taxonomic and probabilistic constraints:

Lemma 3.6 *Let $A, B \in \mathcal{C}$.*

a) $A \xrightarrow{1} B \Longleftrightarrow A \subseteq B$

b) $A \xrightarrow{0} B \Longleftrightarrow A \parallel B$

Example 3.7 (Ex. 3.4 continued) From Fig. 2 we e.g. know that **pilots** $\subseteq$ **employees**, hence **pilots** $\xrightarrow{1}$ **employees** and **pilots** $\underset{0,0.05}{\overset{1}{\longleftrightarrow}}$ **employees** hold. □

An important implication of this simple lemma is that on the one hand taxonomic constraints can be incorporated in the probabilistic deduction process and on the other hand that probabilistic information can have a feedback on taxonomic constraints.

3.3 Modeling-consistency of taxonomic and probabilistic constraints

While taxonomic knowledge-bases are always satisfiable, this generally does not hold for probabilistic knowledge-bases. A probabilistic knowledge-base $\mathcal{P}$ is satisfiable, iff $\mathcal{O} = \emptyset$ is not a logical consequence of $\mathcal{P}$. For the context of OODBs we introduce *modeling-consistency* as an even stronger notion of satisfiability. The definition of modeling-consistent probabilistic knowledge-bases naturally follows from the restriction of probabilistic interpretations as introduced in Sec. 3.2 and the assumption that the extension of every basic class can be different from $\emptyset$ and $\mathcal{O}$ (modeling-consistency of taxonomic knowledge-bases).

Definition 3.8 A probabilistic knowledge-base $\mathcal{T}$ is called *modeling-consistent*, iff there exists a model $(\mathcal{O}, J, \mathrm{P})$ of $\mathcal{P}$ with $0 < \mathrm{P}(J(B_i)) < 1$ for all $B_i \in \mathcal{B}$. □

Since in general a probabilistic knowledge-base may be modeling-inconsistent, it is important to elaborate techniques to check its modeling-consistency. This can be achieved by applying linear programming techniques. It must be checked, if a corresponding set of inequalities is solvable. In the sequel we assume that the definition of taxonomic constraints is extended to general t-class terms.

Definition 3.9 Let $I := \{C_1 \ldots C_k \mid C_i = B_i \text{ or } C_i = \overline{B}_i \text{ with } B_i \in \mathcal{B} \text{ for } i \in [1:k]\}$. For all taxonomic knowledge-bases $\mathcal{T}$ and all probabilistic knowledge-bases $\mathcal{P}$ with $\mathcal{T} \subseteq \mathcal{P}$ and $\mathcal{P} \backslash \mathcal{T} \subseteq \mathcal{PC}$ we define the set of linear inequalities $\mathcal{D}_{\mathcal{T},\mathcal{P}}$ over the variables $\{x_C \mid C \in I, \mathcal{T} \not\models C = \emptyset\}$ by:

a) $C \in I, \mathcal{T} \not\models C = \emptyset \implies \quad x_C \geq 0 \quad \in \mathcal{D}_{\mathcal{T},\mathcal{P}}$,

b) $\sum\{x_C \mid C \in I, \mathcal{T} \not\models C = \emptyset\} = 1 \quad \in \mathcal{D}_{\mathcal{T},\mathcal{P}}$,

c) $A \xrightarrow{u_1,u_2} B \in \mathcal{P}$ or $A \underset{v_1,v_2}{\overset{u_1,u_2}{\longleftrightarrow}} B \in \mathcal{P}$ or
$B \underset{u_1,u_2}{\overset{v_1,v_2}{\longleftrightarrow}} A \in \mathcal{P} \implies$

$$\sum_{C\in I,\mathcal{T}\not\models C=\emptyset,\models C\subseteq A} u_1 x_C \leq \sum_{C\in I,\mathcal{T}\not\models C=\emptyset,\models C\subseteq AB} x_C \quad \in \mathcal{D}_{\mathcal{T},\mathcal{P}} ,$$

$$\sum_{C\in I,\mathcal{T}\not\models C=\emptyset,\models C\subseteq AB} x_C \leq \sum_{C\in I,\mathcal{T}\not\models C=\emptyset,\models C\subseteq A} u_2 x_C \quad \in \mathcal{D}_{\mathcal{T},\mathcal{P}} .$$

□

Theorem 3.10 *Let $\mathcal{T}$ be a taxonomic knowledge-base and $\mathcal{P}$ be a probabilistic knowledge-base with $\mathcal{T} \subseteq \mathcal{P}$ and $\mathcal{P}\backslash\mathcal{T} \subseteq \mathcal{PC}$.*

a) $\mathcal{P}$ is modeling-consistent, iff the following set of linear inequalities is solvable:

$$\mathcal{D}_{\mathcal{T},\mathcal{P}} \cup \{0 < \sum_{C\in I,\mathcal{T}\not\models C=\emptyset,\models C\subseteq B} x_C < 1 \mid B \in \mathcal{B}\} \quad (1)$$

b) $\mathcal{P}$ is modeling-consistent, iff the following sets of linear inequalities are solvable for all $B \in \mathcal{B}$:

$$\mathcal{D}_{\mathcal{T},\mathcal{P}} \cup \{0 < \sum_{C\in I,\mathcal{T}\not\models C=\emptyset,\models C\subseteq B} x_C\} \quad (2)$$

$$\mathcal{D}_{\mathcal{T},\mathcal{P}} \cup \{\sum_{C\in I,\mathcal{T}\not\models C=\emptyset,\models C\subseteq B} x_C < 1\} \quad (3)$$

c) The complexity of the modeling-consistency test is polynomial in the size of $\mathcal{D}_{\mathcal{T},\mathcal{P}}$.

Proof:
a) For similar considerations refer to [ADP91], [NS92] or [CL94].

b) "⇒": the claim directly follows from a).
"⇐": Let $n = |\{C \in I \mid \mathcal{T} \not\models C = \emptyset\}|$. For $i \in [1 : k]$ let $x_{i,0} \in [0,1]^n$ be solutions of the k systems of linear inequalities given by (2). For $i \in [1 : k]$ let $x_{i,1} \in [0,1]^n$ be solutions of the k systems of linear inequalities given by (3). A solution of (1) is given by $x = \frac{1}{2k}(\sum_{i\in[1:k]}(x_{i,0} + x_{i,1}))$.

c) We can prove that the systems of linear inequalities given by (2) and (3) are solvable by maximizing and minimizing $\sum_{C\in I,\mathcal{T}\not\models C=\emptyset,\models C\subseteq B_i} x_C$ for all $B_i \in \mathcal{B}$ subject to $\mathcal{D}_{\mathcal{T},\mathcal{P}}$. The systems of linear inequalities given by (2) and (3) are solvable, iff the linear optimization problems have a solution with a maximum greater than 0 and a minimum less than 1, respectively. Thus the modeling-consistency of a probabilistic knowledge-base can be checked by solving $2k$ linear optimization problems, each in polynomial time in the number of variables and the number of constraints (see e.g. [PS82]). □

The modeling-consistency test as described in the proof of Theorem 3.10 c) is illustrated by the following example.

Example 3.11 Let $\mathcal{A} = \{\emptyset, \mathcal{O}, \mathsf{A}, \mathsf{B}\}$, $\mathcal{T} = \{\mathsf{A} \parallel \mathsf{B} = \mathcal{O}\}$, $\mathcal{P} = \mathcal{T} \cup \{\mathcal{O} \xrightarrow{0.2,0.4} \mathsf{A}, \mathcal{O} \xrightarrow{0.6,0.7} \mathsf{B}\}$.

The set of inequalities $\mathcal{D}_{\mathcal{T},\mathcal{P}}$ is given by:

$$\begin{array}{rcccl} 0.2 \cdot (x_{\mathsf{A}\bar{\mathsf{B}}} + x_{\bar{\mathsf{A}}\mathsf{B}}) & \leq & x_{\mathsf{A}\bar{\mathsf{B}}} & \leq & 0.4 \cdot (x_{\mathsf{A}\bar{\mathsf{B}}} + x_{\bar{\mathsf{A}}\mathsf{B}}) \\ 0.6 \cdot (x_{\mathsf{A}\bar{\mathsf{B}}} + x_{\bar{\mathsf{A}}\mathsf{B}}) & \leq & x_{\bar{\mathsf{A}}\mathsf{B}} & \leq & 0.7 \cdot (x_{\mathsf{A}\bar{\mathsf{B}}} + x_{\bar{\mathsf{A}}\mathsf{B}}) \end{array}$$

$$x_{\mathsf{A}\bar{\mathsf{B}}}, x_{\bar{\mathsf{A}}\mathsf{B}} \geq 0$$

$$x_{\mathsf{A}\bar{\mathsf{B}}} + x_{\bar{\mathsf{A}}\mathsf{B}} = 1$$

We get max $x_{\mathsf{A}\bar{\mathsf{B}}} = 0.4$, min $x_{\mathsf{A}\bar{\mathsf{B}}} = 0.3$, max $x_{\bar{\mathsf{A}}\mathsf{B}} = 0.7$ and min $x_{\bar{\mathsf{A}}\mathsf{B}} = 0.6$. Thus $\mathcal{P}$ is modeling-consistent. A solution of (1) is given by:

$$(x_{\mathsf{A}\bar{\mathsf{B}}}, x_{\bar{\mathsf{A}}\mathsf{B}}) = \frac{1}{4}((0.4, 0.6) + (0.3, 0.7) + (0.3, 0.7) + (0.4, 0.6)) .$$

□

In the same way it can be proved that the probabilistic knowledge-base of Ex. 3.4 is modeling-consistent (we get a system of linear inequalities with 24 variables).

Note that for the special case of basic t-classes organized in a partition hierarchy the number of variables is equal to the number of basic t-classes in the lowest level of the hierarchy.[3]

3.4 Checking the taxonomic and probabilistic constraints

Once a modeling-consistent probabilistic knowledge-base is established, we can easily check the integrity of an OODB instance. The integrity of an OODB instance with respect to a taxonomic knowledge-base can simply be checked by testing set-inclusions at runtime. The integrity of an OODB instance with respect to a set of probabilistic constraints can be checked by evaluating the probability measure for each conjunctive t-class term which occurs in an uncertain or correlation rule. Taking e.g. the relative cardinality of classes as probabilistic interpretation, an OODB instance satisfies the uncertain rule $A \xrightarrow{x_1,x_2} B$, iff $A \neq \emptyset$ implies that the proportion of the number of objects in AB to the number of objects in A is contained in the interval $[x_1, x_2]$.

4 Portfolio management application

The TOP database model supports applications which can be characterized by the following criteria:

1) Constraints on the composition of sets must be expressed.

2) The universe can be described by one or more hierarchies.

3) Constraints can be represented by uncertain rules.

4) Constraints occur between different hierarchy levels.

5) Constraints describe a range of values.

Following these criteria, a more complex application of the TOP database model can be provided within the stock market field for the administration of stock funds ([Kra95]). A stock fund consists of different stocks which can be classified according to economically important criteria. This classification determines the taxonomic hierarchy (see the EER-diagram in Fig. 3).

[3] e.g. for $\mathcal{B} = \{\mathsf{B}_{i,j} \mid i \in [1:m], j \in [1:\mu_i]\} \cup \{\mathsf{B}_i \mid i \in [1:m]\}$ and $\mathcal{T} = \{\mathsf{B}_{i,1} \parallel \ldots \parallel \mathsf{B}_{i,\mu_i} = \mathsf{B}_i \mid i \in [1:m]\} \cup \{\mathsf{B}_1 \parallel \ldots \parallel \mathsf{B}_m = \mathcal{O}\}$ with $m \geq 1$ and $\mu_i \geq 1$ for $i \in [1:m]$ we get $\sum_{i\in[1:m]} \mu_i$ variables.

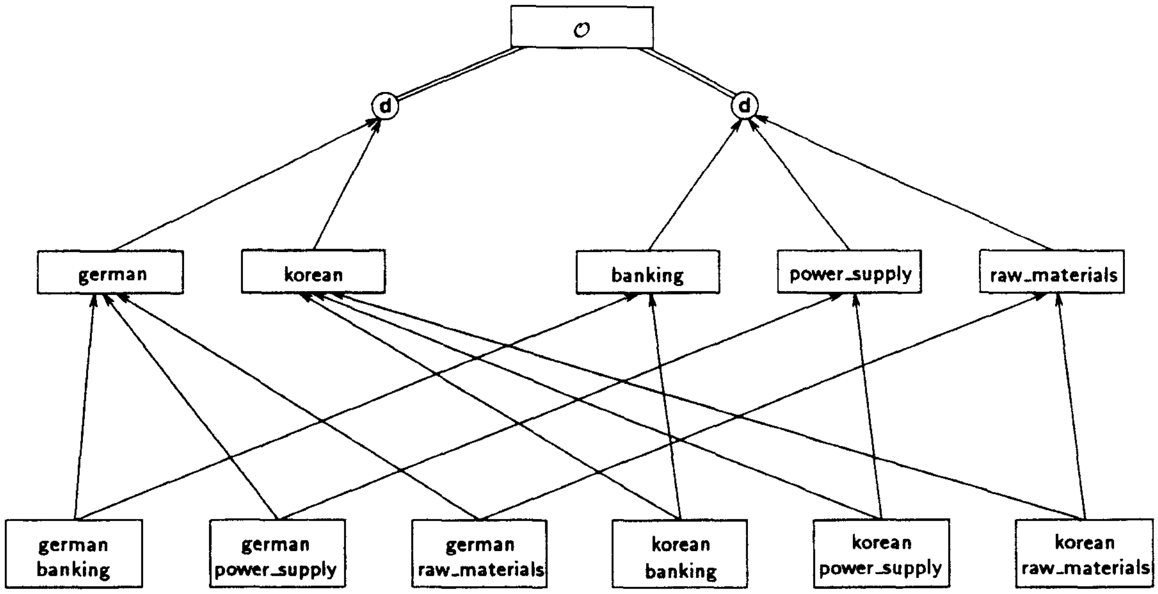

Figure 3: TOP-modeling of a portfolio management application

The decision of the management staff of a stock fund to invest between 30 and 35 per cent of the available capital in banking stocks, to invest between 10 and 40 per cent in power supply stocks and to invest between 40 and 60 per cent in raw material stocks can be expressed by the following probabilistic constraints:

$$\mathcal{O} \xrightarrow{0.3,0.35} \text{banking},$$
$$\mathcal{O} \xrightarrow{0.1,0.4} \text{power\_supply},$$
$$\mathcal{O} \xrightarrow{0.4,0.6} \text{raw\_material}.$$

The composition of a stock fund influences the expected gain and the risk of an investment. Fig. 4 shows the expected gain μ and the risk σ for a stock fund of raw material stocks with respect to different compositions by German and Korean raw material stocks. The expected gain μ and the risk σ can be determined by applying standard methods from the field of economic analysis. For the German and Korean raw material stocks we assume the expected gains $\mu_G = 0.07$ and $\mu_K = 0.085$ and the risks $\sigma_G = 0.052$ and $\sigma_K = 0.107$. The value x denotes the portion of investment which is made with respect to the German raw material stocks.

Uncertain constraints on the composition of classes enable us to guarantee an upper bound for the risk and a lower bound for the expected gain of a stock fund. The maximal risk of 0.05 for German and Korean raw material stocks can be guaranteed by the following probabilistic constraints:

$$\text{raw\_material} \xrightarrow{0.59,0.97} \text{german raw\_material},$$
$$\text{raw\_material} \xrightarrow{0.03,0.41} \text{korean raw\_material}.$$

The minimal expected gain of 0.075 for German and Korean raw material stocks can be guaranteed by:

$$\text{raw\_material} \xrightarrow{0,0.67} \text{german raw\_material},$$
$$\text{raw\_material} \xrightarrow{0.33,1} \text{korean raw\_material}.$$

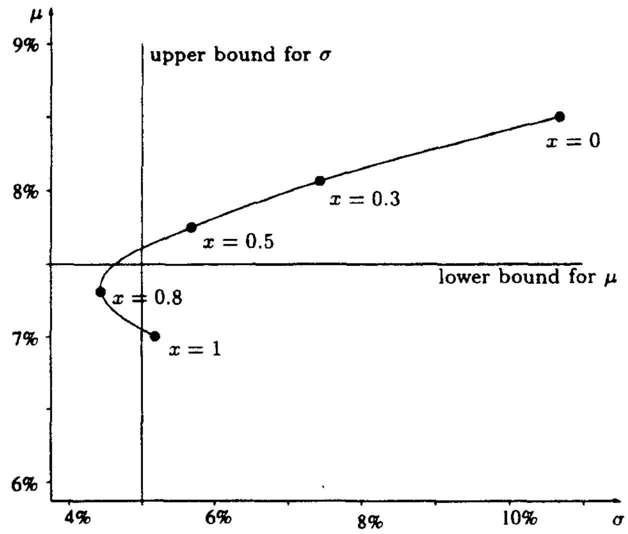

Figure 4: Expected gain and risk of a stock fund

5 Deduction of uncertain integrity constraints

In addition to the crucial problem of proving the modeling-consistency of a probabilistic knowledge-base, it is interesting to derive taxonomic and uncertain integrity constraints from a probabilistic knowledge-base. Note that due to the strong interconnections between taxonomic and uncertain integrity constraints, it might be sometimes difficult to judge the effects that uncertain rules can have on each other. Therefore it is helpful to support the modeling process of a probabilistic knowledge-base for an OODB by deduction techniques for taxonomic and uncertain knowledge. We consider an approach based on local inference rules which naturally supports explanation tools. In this paper we do not explore the complex optimization problem for probabilistic queries in full depth. We just want to highlight the advantages gained by the careful interaction between taxonomic and probabilistic knowledge.

5.1 Internal representation

In this section we present an internal representation of t-classes which enables us to reduce the search space in probabilistic deduction and to evaluate taxonomic relationships in the premise of inference rules. For the special case of taxonomic knowledge-bases without partition constraints this internal representation of t-classes enables us to perform taxonomic reasoning in linear time in the size of the taxonomic knowledge-base. Since there are different t-class terms with the same interpretation due to the axioms of set theory (e.g. $AA = A$, $A\emptyset = \emptyset$, ...) or because of taxonomic constraints (e.g. $A \subseteq B \implies AB = A$), we want to identify t-class terms with an identical interpretation. For this purpose we define an equivalence relation that takes into account all set-theoretic laws and taxonomic integrity constraints.

Definition 5.1 Let $\mathcal{T}$ be a taxonomic knowledge-base and $\mathcal{G}$ be the set of all general t-class terms.

a) The equivalence relation $\sim_{\mathcal{T}}$ on $\mathcal{G}$ is defined by:

$$G_1 \sim_{\mathcal{T}} G_2 \quad :\Leftrightarrow \quad \mathcal{T} \models G_1 = G_2 .$$

b) For $G \in \mathcal{G}$ let $[G]_{\sim_{\mathcal{T}}}$ be abbreviated by $G_{\mathcal{T}}$. Let $\mathcal{B}_{\mathcal{T}} := \{B_{\mathcal{T}} \mid B \in \mathcal{B}\}$, $\mathcal{C}_{\mathcal{T}} := \{C_{\mathcal{T}} \mid C \in \mathcal{C}\}$ and $\mathcal{G}_{\mathcal{T}} := \{G_{\mathcal{T}} \mid G \in \mathcal{G}\}$.

c) The partial order $\subseteq$ is canonically extended to $\mathcal{G}_{\mathcal{T}}$ by:[4]

$$A_{\mathcal{T}} \subseteq_{\mathcal{T}} B_{\mathcal{T}} \quad :\Leftrightarrow \quad (AB)_{\mathcal{T}} = A_{\mathcal{T}} .$$

d) The operations $\cap$, $\cup$ and $^{-}$ are canonically extended to $\mathcal{G}_{\mathcal{T}}$ by:

$$\begin{array}{lcl} A_{\mathcal{T}} B_{\mathcal{T}} & := & (AB)_{\mathcal{T}}, \\ A_{\mathcal{T}} \cup_{\mathcal{T}} B_{\mathcal{T}} & := & (A \cup B)_{\mathcal{T}}, \\ \mathrm{comp}_{\mathcal{T}}(A_{\mathcal{T}}) & := & (\overline{A})_{\mathcal{T}}. \end{array}$$

□

All conjunctive t-class terms are assumed to be represented by their corresponding elements of $\mathcal{C}_{\mathcal{T}}$. This presumes that taxonomic and probabilistic constraints are defined on $\mathcal{C}_{\mathcal{T}}$. Taxonomic constraints are extended to $\mathcal{C}_{\mathcal{T}}$ by Def. 5.1, probabilistic constraints can similarly be extended to $\mathcal{C}_{\mathcal{T}}$. The translation of the user-defined probabilistic constraints over $\mathcal{C}$ into the corresponding probabilistic constraints over $\mathcal{C}_{\mathcal{T}}$ is done as follows:

(IR) Internal Representation:

$$\text{(a)} \ \{ A \xrightarrow{x_1,x_2} B \} \vdash A_{\mathcal{T}} \xrightarrow{x_1,x_2} A_{\mathcal{T}} B_{\mathcal{T}}$$

$$\text{(b)} \ \{ A \underset{y_1,y_2}{\overset{x_1,x_2}{\longleftrightarrow}} B \} \vdash A_{\mathcal{T}} \underset{y_1,y_2}{\overset{x_1,x_2}{\longleftrightarrow}} B_{\mathcal{T}}$$[5]

[4]The definitions in c) and d) are independent from the representative of the equivalence classes.

[5]$A_{\mathcal{T}} \underset{y_1,y_2}{\overset{x_1,x_2}{\longleftrightarrow}} B_{\mathcal{T}}$ is equivalent to $A_{\mathcal{T}} \xrightarrow{x_1,x_2} A_{\mathcal{T}} B_{\mathcal{T}}$ and $B_{\mathcal{T}} \xrightarrow{y_1,y_2} A_{\mathcal{T}} B_{\mathcal{T}}$.

The internal representation over $\mathcal{C}_{\mathcal{T}}$ yields an enormous search space reduction that can be illustrated by the following example.

Example 5.2 Let the alphabet $\mathcal{A}$ be defined by $\mathcal{A} = \{\emptyset, \mathcal{O}, \mathsf{A}, \mathsf{B}, \mathsf{C}, \mathsf{D}\}$ and let the taxonomic knowledge-bases $\mathcal{T}_0$ and $\mathcal{T}_1$ be given by:

$$\mathcal{T}_0 \ = \ \emptyset, \quad \mathcal{T}_1 \ = \ \{\mathsf{A} \cup \mathsf{B} \subseteq \mathsf{C}, \mathsf{A} \parallel \mathsf{B}, \mathsf{D} \subseteq \mathsf{A}\} \ .$$

The following table gives a comparison of the number of elements in $\mathcal{C}_{\mathcal{T}}$, the number of uncertain rules and the number of correlation rules over $\mathcal{C}_{\mathcal{T}}$ occurring w.r.t. $\mathcal{T} = \mathcal{T}_0$ and $\mathcal{T} = \mathcal{T}_1$.

	$\mathcal{T} = \mathcal{T}_0$	$\mathcal{T} = \mathcal{T}_1$
number of elements in $\mathcal{C}_{\mathcal{T}}$	17	6
number of uncertain rules over $\mathcal{C}_{\mathcal{T}}$	82	19
number of correlation rules over $\mathcal{C}_{\mathcal{T}}$	289	36

□

Note that for the special case of basic t-classes organized in a partition hierarchy, the number of elements in $\mathcal{C}_{\mathcal{T}}$ is equal to the number of elements in the alphabet.

5.2 Probabilistic inference rules

As already pointed out, the deduction of probabilistic knowledge is assumed to support the design of a set of modeling-consistent probabilistic constraints. During this process the user may ask the uncertain query $?A \xrightarrow{x_1,x_2} B$. If $A = \emptyset$ holds, the uncertain rule $?A \xrightarrow{x_1,x_2} B$ is always true. If $A \subseteq B$ or $A \parallel B$ hold, the answer $x_1 = x_2 = 1$ or $x_1 = x_2 = 0$, resp., can be returned by taxonomic deduction without engaging in probabilistic deduction. Otherwise probabilistic query evaluation has to be initiated, employing probabilistic inference rules.

Below we state an inference rule for the chaining of correlation rules. As proved in [Luk95] this inference rule is sound and yields the tightest bounds for taxonomic knowledge-bases without partition constraints. Note that taxonomic constraints (to be evaluated on the internal representation of t-classes) appear in the premise of the inference rules.

Let $A, B, C \in \mathcal{C}$ and $\mathcal{T}$ be a taxonomic knowledge-base with $\mathcal{T} \not\models A \subseteq C$ and $\mathcal{T} \not\models AC = \emptyset$. Let $\not\subseteq_{\mathcal{T}} := \mathcal{G}_{\mathcal{T}} \times \mathcal{G}_{\mathcal{T}} \setminus \subseteq_{\mathcal{T}}$.

(CH) Chaining of correlation rules:

$\{ A_{\mathcal{T}} \underset{v_1,v_2}{\overset{u_1,u_2}{\longleftrightarrow}} B_{\mathcal{T}}, B_{\mathcal{T}} \underset{y_1,y_2}{\overset{x_1,x_2}{\longleftrightarrow}} C_{\mathcal{T}}, u_1, v_1, x_1, y_1 \ > \ 0 \} \vdash$ $A_{\mathcal{T}} \xrightarrow{z_1,z_2} C_{\mathcal{T}}$ with z_1 equal to

$$\begin{cases} \frac{u_1}{y_2} & \text{if } C_{\mathcal{T}} \subseteq_{\mathcal{T}} A_{\mathcal{T}}, A_{\mathcal{T}} B_{\mathcal{T}} \subseteq_{\mathcal{T}} C_{\mathcal{T}} \\ \frac{u_1 x_1}{v_2 y_2} & \text{if } C_{\mathcal{T}} \subseteq_{\mathcal{T}} A_{\mathcal{T}}, A_{\mathcal{T}} B_{\mathcal{T}} \not\subseteq_{\mathcal{T}} C_{\mathcal{T}} \\ u_1 & \text{if } C_{\mathcal{T}} \not\subseteq_{\mathcal{T}} A_{\mathcal{T}}, A_{\mathcal{T}} B_{\mathcal{T}} \subseteq_{\mathcal{T}} C_{\mathcal{T}} \\ \frac{u_1 x_1}{v_2} & \text{if } C_{\mathcal{T}} \not\subseteq_{\mathcal{T}} A_{\mathcal{T}}, A_{\mathcal{T}} B_{\mathcal{T}} \not\subseteq_{\mathcal{T}} C_{\mathcal{T}}, \\ & \quad B_{\mathcal{T}} C_{\mathcal{T}} \subseteq_{\mathcal{T}} A_{\mathcal{T}} \\ \max(0, u_1 - \frac{u_1}{v_1} + \frac{u_1 x_1}{v_1}) & \text{otherwise} \end{cases}$$

and z_2 equal to

$$
\begin{cases}
\min(1-u_1, \frac{u_2(1-y_1)\min(x_2,1-v_1)}{v_1 y_1}, & \\
\quad \frac{(1-y_1)\min(x_2,1-v_1)}{y_1 v_1+(1-y_1)\min(x_2,1-v_1)}) & \text{if } A_T B_T C_T \subseteq_T \emptyset_T \\
\min(u_2, \frac{u_2 x_2}{v_1}) & \text{if } A_T B_T C_T \not\subseteq_T \emptyset_T, \\
& \quad A_T C_T \subseteq_T B_T \\
\min(1, \frac{u_2 x_2}{v_1 y_1}, 1-u_1+\frac{u_1 x_2}{v_1}, & \\
\quad \frac{u_2}{y_1}, \frac{x_2}{v_1 y_1+(1-y_1)x_2}) & \text{if } A_T B_T C_T \not\subseteq_T \emptyset_T, \\
& \quad A_T C_T \not\subseteq_T B_T, \\
& \quad B_T C_T \subseteq_T A_T \\
\min(1, \frac{u_2 x_2}{v_1 y_1}, 1-u_1+\frac{u_1 x_2}{v_1}, & \\
\quad u_2-\frac{u_2 x_2}{v_1}+\frac{u_2 x_2}{v_1 y_1}, & \\
\quad \frac{x_2}{y_1 v_1+(1-y_1)x_2}) & \text{otherwise}
\end{cases}
$$

Example 5.3 With respect to the portfolio management application we could draw the following conclusions by the chaining of correlation rules. Since german raw_material $\subseteq \mathcal{O}$, raw_material $\not\subseteq$ german raw_material (second case for the lower bound) and german raw_material $\not\subseteq \emptyset$, german raw_material $\subseteq$ raw_material (second case for the upper bound), we get:

$$
\{\mathcal{O} \xleftrightarrow[1]{0.4,0.6} \text{raw\_material}, \quad \text{raw\_material} \xleftrightarrow[1]{0.59,0.67} \text{german raw\_material}\} \vdash_{(CH)} \mathcal{O} \xrightarrow{0.236,0.402} \text{german raw\_material}
$$

Since korean raw_material $\subseteq \mathcal{O}$, raw_material $\not\subseteq$ korean raw_material (second case for the lower bound) and korean raw_material $\not\subseteq \emptyset$, korean raw_material $\subseteq$ raw_material (second case for the upper bound), we get:

$$
\{\mathcal{O} \xleftrightarrow[1]{0.4,0.6} \text{raw\_material}, \quad \text{raw\_material} \xleftrightarrow[1]{0.33,0.41} \text{korean raw\_material}\} \vdash_{(CH)} \mathcal{O} \xrightarrow{0.132,0.246} \text{korean raw\_material}
$$

Hence the probabilistic constraints for the portfolio management application restrict the investments into German raw material stocks to the range of 23.6 to 40.2 per cent and the investments into Korean raw material stocks to the range of 13.2 to 24.6 per cent. The user can check the deduced uncertain rules against his intentions and possibly change the constraints.

6 Related work

We are not aware of any work dealing with uncertain integrity constraints as a generalization of taxonomic hierarchies in object-oriented databases. Some related ideas are provided by absolute cardinality constraints considered in the context of ER-modeling.

• Calvanese and Lenzerini [CL94] examine the interaction between ISA-relationships and cardinality constraints. They show that the satisfiability of a single class in an ER-model with ISA-relationships and cardinality constraints can be checked by solving a corresponding linear programming problem.

The combination of taxonomic and uncertain knowledge was also examined by Ng and Subrahmanian [NS92].

• Ng and Subrahmanian present an approach to integrate empirical probabilities in deductive databases. An empirical program consists of two parts, true/false knowledge about classes of individuals (or single individuals) and empirical clauses representing statistical knowledge about "generic" individuals. In a compilation step this knowledge base is enriched by adding logically entailed first order and empirical clauses. The consistency of an empirical program is checked by integer linear programming. Queries about individuals are either answered by deduction or by induction on the enriched knowledge base. In contrast to this approach in which *integer* linear programming techniques are used to verify the consistency of a knowledge-base, we showed that within our framework the modeling-consistency of a probabilistic knowledge-base can be proved by general linear programming.[6] Furthermore we apply more general and more precise inference rules on the uncertain rule knowledge allowing a broader range of hypothetical reasoning. Here we do not consider queries considering individuals, but our approach can be extended to uncertain facts as well ([TKG95]).

7 Conclusion and Outlook

We have presented the TOP database model as a coherent and evolutionary approach to extend current OODB technology by taxonomic and uncertain modeling and reasoning capabilities. For the complex field of uncertain deduction we applied novel techniques and algorithms capable of exploiting taxonomic knowledge during the probabilistic deduction process. We expect applications such as configuration tasks, multimedia, inventory control, lead qualification or other management tasks under uncertain constraints to be suitable for a TOP database system. In this paper we have only examined *hard probabilistic constraints*, i.e., probabilistic constraints which must be satisfied by each class extension. A more general scenario of how the TOP-features support interesting application domains might be as follows. As a generalization *soft probabilistic constraints* representing intended or desired restrictions, i.e., restrictions that may be violated by a class extension are essential. The portfolio management application illustrates the usage of these notions: regulations about the composition of growth funds according to the economic law or sales prospectus are examples for mandatory probabilistic constraints. Additionally the brokers may desire that the funds contain a certain percentage of computer industry stock which is an example of an optional probabilistic constraint. The database system should strive after the satisfaction of this type of restrictions in the long run where the level of satisfaction is determined by an evaluation function. Violations, however, especially short term ones are possible. The system could even try to satisfy the restrictions by automatically triggering update actions.

Of course there is more research to be done, e.g. extending TOP to cover the full ODMG spectrum including relationships. More work has also to be done with respect to enriching the TOP database model by the capability to represent uncertain knowledge about individual objects and the attributes of objects. A project to build a TOP-prototype is under way using available OODBs like O_2 or Versant. The implementation of uncertain deduction can adapt our prior experiences with the DUCK-system.

[6] Note that the time complexity of general linear programming is polynomial in the number of constraints and the number of variables, while integer linear programming is known to be in $\mathcal{NP}$ (see e.g. [PS82]).

8 Acknowledgements

We are grateful to Peter Kratz for his contributions with respect to the portfolio management application. We are also grateful to Martin Russling for fruitful discussions on economic applications of our approach.

References

[ADP91] Stéphane Amarger, Didier Dubois, and Henri Prade. Constraint propagation with imprecise conditional probabilities. In Bruce D. D'Ambrosio, Philippe Smets, and Piero P. Bonissone, editors, *Proc. of the 7 th Conference on Uncertainty in Artificial Intelligence*, pages 26–34, Los Angeles, CA, 1991. Morgan Kaufmann Publishers.

[BGMP92] Daniel Barbará, Hector Garcia-Molina, and Daryl Porter. The management of probabilistic data. *IEEE Transactions on Knowledge and Data Engineering*, 4(5):487–502, Oct. 1992.

[Bra91] Ronald J. Brachman. *Knowledge representation*. MIT Press, 1991.

[Cat94] R. G. G. Cattel, editor. *The Object Database Standard: ODMG-93*. Morgan Kaufmann Publishers, San Mateo, California, 1994.

[CL94] Diego Calvanese and Maurizio Lenzerini. Making object-oriented schemas more expressive. In *Proc. ACM SIGACT-SIGMOD Symp. on Principles of Database Systems*, pages 243–254, Minnesota, USA, May 1994.

[GKT91] Ulrich Güntzer, Werner Kießling, and Helmut Thöne. New directions for uncertainty reasoning in deductive databases. In *Proc. ACM SIGMOD Conference*, pages 178–187, Denver, CO, May 1991.

[KLKG94] Werner Kießling, Thomas Lukasiewicz, Gerhard Köstler, and Ulrich Güntzer. The TOP database model – taxonomy, object-orientation and probability. In *Proc. Int'l Workshop on Uncertainty in Databases and Deductive Systems*, pages 71–82, Ithaca, New York, Nov. 1994.

[Kra95] Peter Kratz. Betriebswirtschaftliche Anwendungen im TOP-Datenbankmodell (Economic applications in the TOP database model). Master's thesis, Universität Augsburg, 1995.

[LS94] V.S. Lakshmanan and F. Sadri. Modeling uncertainty in deductive databases. In Dimitris Karagiannis, editor, *Proc. 5th International Conference on Database and Expert Systems Applications*, pages 724–733, Athens, Greece, Sept. 1994.

[Luk95] Thomas Lukasiewicz. Uncertain reasoning in concept lattices. In *Proc. of the 3 rd European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU)*, volume 946 of *Lecture Notes in Artificial Intelligence*, pages 293–300, 1995.

[NS92] R. T. Ng and V. S. Subrahmanian. Empirical probabilities in monadic deductive databases. In Didier Dubois, Michael P. Wellman, Bruce D' Ambrosio, and Phillipe Smets, editors, *Proc. of the 8 th Conference on Uncertainty in Artificial Intelligence*, pages 215–222, Stanford, CA, Jul. 1992. Morgan Kaufmann Publishers.

[Pea88] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, 1988.

[PS82] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization, Algorithms and Complexity*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982.

[Som90] Léa Sombé. Reasoning under incomplete information in artificial intelligence: A comparison of formalisms using a single example. *International Journal of Intelligent Systems*, 5(4):323–472, 1990.

[TGK95] Helmut Thöne, Ulrich Güntzer, and Werner Kießling. Modeling, chaining and fusion of uncertain knowledge. In Tok Wang Ling and Yoshifumi Masunaga, editors, *Proc. of the 4 th Int'l. Conference on Database Systems for Advanced Applications*, pages 197–205, April 1995.

[Thö94] Helmut Thöne. *Precise Conclusion under Uncertainty and Incompleteness in Deductive Database Systems*. PhD thesis, Universität Tübingen, July 1994.

[TKG95] Helmut Thöne, Werner Kießling, and Ulrich Güntzer. On cautious probabilistic inference and default detachment. *Special issue of the Annals of Operations Research*, 55:195–224, Scientific Publ. Comp., Netherlands, 1995.