Multi-Interpretation Operators and Approximate Classification

Joeri Engelfriet, Jan Treur

Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081, 1081 HV Amsterdam
Email: {joeri, treur}@cs.vu.nl
URL: http://www.cs.vu.nl/~{joeri, treur}

Abstract
In this paper non-classical logical techniques are introduced to formalize the analysis of multi-interpretable observation information, in particular in approximate classification processes where information on attributes of an object is to be inferred on the basis of observable properties of the object. One frequently occurring reason for imperfect classification is when the available observations are insufficient to determine unique values for each of the attributes: a range of values may still be possible. Another often occurring reason for imperfect classification occurs when the observation information is contradictory: for some of the attributes not any value is possible. The combination of both types of imperfection is non-trivial from a standard logical perspective. To address this problem multi-interpretation operators and selection operators are introduced; these techniques generalize non-monotonic reasoning formalisms such as default logic. A specific multi-interpretation operator for approximate classification is introduced and formally analysed. On the basis of this approach, in co-operation with industry a system has been designed and implemented for the analysis of ecological monitoring information.

Keywords: approximate classification, belief sets, nonmonotonic, interpretation

1 Introduction

In most real-life situations humans receive information that can be interpreted in many different ways. On the one hand this involves interpretation: the information from the outside world has to be given a meaning. In logic the notion of interpretation mapping has been introduced to describe the interpretation of one logical theory in another logical theory, for example geometry in algebra (cf. Chapter 5 in [7]). This notion assumes a choice for one interpretation, and does not cover cases in which multiple interpretations at the same time are relevant. But on the other hand, given the information received, there is often more than one possibility for forming a set of beliefs about the world. This can be due to incompleteness, vagueness or uncertainty of the input, and may require non-monotonic reasoning techniques of the reasoning agent. The context often determines the view with which this information is interpreted. In this paper multi-interpretation operators are introduced and applied to formalize multiple interpretations of observation information. The notion of a multi-interpretation operator is rather general: it subsumes on the one hand the notion of interpretation in logic, and on the other hand the notion of (non-monotonic) belief set operator as introduced in [5].

A specific type of multi-interpretation operator is defined to interpret observation information in approximate classification tasks. The generic task formalized by such an operator is as follows. Suppose there is an object in the world, and one is interested in the values of attributes of this object. It is possible to observe the object leading to input information consisting of observable properties. On the basis of these properties information on the values of attributes of the object is derived. This task involves interpretation: interpreting observable properties in terms of values of attributes (which may be difficult or impossible to observe directly).

Two problems occurring often in such classification tasks in real-world domains are underspecification and overspecification. Underspecification occurs when the observations are sufficient to exclude some of the values of attributes, but insufficient to determine unique values for each of the attributes: a range of values may still be possible. Overspecification occurs when the observation information is contradictory: for some of the attributes not any value is possible. Underspecification can lead to an approximation (an upper bound) of the solution of the classification: a set of possibilities, one of which is the right solution. If the number of observations increases, the approximation may come closer to a unique solution: the resulting sets of possible classifications will decrease with the increase of observation information. Overspecification leads to a trivial approximation from the other direction: the empty set as a lower bound (no classification at all). The combination of underspecification and overspecification as occurs often in practical domains is problematic. The occurrence of contradictory observation information interferes with the approximations that can be used as upper bound of the solution.

Multi-interpretation operators can be used to clarify this interference: such an operator formalizes that there is more than one possibility of interpreting the observed findings. A generic multi-interpretation operator is introduced to formalize such tasks. The input language of the operator is restricted to observation information only; interpretations of this observation information are expressed in terms of the output language of the operator. This formalization identifies and separates the overspecification and underspecification and entails an approximate solution of a classification problem in the form of multiple approximations.

One domain in which multi-interpretable observations can be analysed using a technique based on the distinction of different views, is the domain of ecology. Here the possible values of abiotic factors such as moisture and acidity of a terrain, are determined on the basis of the plant species found on the terrain.

The structure of this paper is as follows. Section 2 introduces multi-interpretation operators and selection operators and some properties they may have. The generic multi-interpretation operator for interpreting observation information is also given and studied in this section. The application of these techniques in the domain of ecology is briefly sketched in Section 3. In Section 4, it is shown that such a generic operator is representable in default logic. The last section contains the conclusions.

2 Multi-interpretation Operators and Approximate Classification

In this section the notion of multi-interpretation operator is introduced (Section 2.1), a specific type of multi-interpretation operator is defined that formalizes approximate classification (Section 2.2), and some properties of this multi-interpretation operator are proven (Section 2.3).

2.1 Multi-interpretation operators

A multi-interpretation operator is an operator that assigns to each set of input information, a set of interpretations. The input information is described by propositional formulae in a propositional language $\mathbf{L_1}$. An interpretation is a set of propositional formulae, which is closed under the standard propositional consequence operator **Cn**. Such a closed set will be called a *belief set*, and we assume that they are based on a (possibly different) propositional language $\mathbf{L_2}$. A belief set can be seen as a possible set of beliefs of an agent with perfect (propositional) reasoning capabilities.

Definition 2.1 (Multi-interpretation operator)

a) A *multi-interpretation operator* **MI** with input language $\mathbf{L_1}$ and output language $\mathbf{L_2}$ is a function $\mathbf{MI} : \wp(\mathbf{L_1}) \rightarrow \wp(\wp(\mathbf{L_2}))$ that assigns a set of belief sets to each set of input facts.

b) A multi-interpretation operator **MI** satisfies *non-inclusiveness* if for all $\mathbf{X} \subseteq \mathbf{L_1}$ and all $\mathbf{S}, \mathbf{T} \in \mathbf{MI(X)}$, if $\mathbf{S} \subseteq \mathbf{T}$ then $\mathbf{S} = \mathbf{T}$.

c) The *kernel* $\mathbf{K_{MI}} : \wp(\mathbf{L_1}) \rightarrow \wp(\mathbf{L_2})$ of **MI** is defined by: for all $\mathbf{X} \subseteq \mathbf{L_1}$

$$\mathbf{K_{MI}(X)} = \bigcap \mathbf{MI(X)}.$$

d) If $\mathbf{L_1} \subseteq \mathbf{L_2}$, then a multi-interpretation operator **MI** satisfies *inclusion* if for all $\mathbf{X} \subseteq \mathbf{L_1}$ and all $\mathbf{T} \in \mathbf{MI(X)}$ it holds $\mathbf{X} \subseteq \mathbf{T}$.

The condition of non-inclusiveness guarantees a relative maximality of the possible interpretations. The kernel of a multi-interpretation operator yields the most certain conclusions given a set of initial facts, namely those which are in every possible interpretation of the input information. The last condition expresses conservativity: it means that a possible interpretation of the world at least satisfies the given facts; in this case the multi-interpretation operator defines a method of extending partial information. Note that when **MI(X)** has exactly one element this means that the set $\mathbf{X} \subseteq \mathbf{L_1}$ has a unique interpretation under **MI**.

To give an example of a multi-interpretation operator, consider a set of default rules (the reader is referred to the next section for a definition of default logic). A set of initial facts, together with the default rules, gives rise to a number of extensions (which can be considered belief sets). An operator that assigns the corresponding set of extensions to each set of initial facts is a multi-interpretation operator. The kernel of this operator yields the sceptical (see e.g., [10]) conclusions.

Often, after a number of belief sets have been generated, the reasoning agent will focus on (or make a commitment to) one (or possibly more) of the belief sets, because it seems the most promising, or interesting, possible view on the world. This selection process can be formalized by selection operators (see [5]).

Definition 2.2 (Selection operator and selective interpretation operator)

a) A *selection operator* **s** is a function $\mathbf{s} : \wp(\wp(\mathbf{L})) \rightarrow \wp(\wp(\mathbf{L}))$ that assigns to each set of belief sets a subset (for all $\mathbf{A} \subseteq \wp(\mathbf{L})$ it holds $\mathbf{s(A)} \subseteq \mathbf{A}$) such that whenever $\mathbf{A} \subseteq \wp(\mathbf{L})$ is non-empty, **s(A)** is non-empty. A selection operator **s** is *single-valued* if for all non-empty **A** the set **s(A)** contains exactly one element.

b) A *selective interpretation operator for* the multi-interpretation operator **MI** is a function $\mathbf{C} : \wp(\mathbf{L_1}) \rightarrow \wp(\mathbf{L_2})$ that assigns a belief set to each set of facts, such that for all $\mathbf{X} \subseteq \mathbf{L_1}$ it holds $\mathbf{C(X)} \in \mathbf{MI(X)}$

It is straightforward to check that if $\mathbf{s} : \wp(\wp(\mathbf{L_2})) \rightarrow \wp(\wp(\mathbf{L_2}))$ is a single-valued selection operator, then a selective interpretation operator **C** for a multi-interpretation operator **MI** can be defined by setting

$\mathbf{C(X) = s(MI(X))}$ for all $\mathbf{X} \subseteq \mathbf{L_1}$.

The type of operator described above is very general, and many forms of reasoning can be captured with it (see [5] for a number of examples described in terms of belief set operators). Below, we will describe a generic type of operator applicable for a specific classification task.

2.2 A Multi-interpretation Operator for Approximate Classification

Suppose we have an object in the real world (a car, for example), and we are interested in the values of certain attributes of this object (such as the amount of horsepower of the engine; we assume attributes are functions). All we can do is observe a number of properties of the object (such as the colour, or maybe that it is a Ford). Knowledge relating observable properties to the possible values of attributes is needed to perform this classification task. Using this knowledge, for each attribute certain values can be excluded. In a situation of underspecification for each of the attributes this results in a remaining range of possible values. However, if also overspecification occurs, then in a classical manner it can be derived that for a certain attribute no value at all is possible, which contradicts the functional nature of attributes.

A formalization of this approximate classification task can be made using the notions defined above. The language $\mathbf{L_1}$ is the propositional language of which the atoms are the ground atoms defined by the following signature:

a finite set **Props** of property names:	$\mathbf{p_1, ..., p_k}$
a unary predicate:	**observed**

The meaning of $\mathbf{observed(p_i)}$ is (not surprisingly) that the property $\mathbf{p_i}$ has been observed of the object. A variable over the set **Props** will be denoted by **P**.

The language $\mathbf{L_2}$ is the propositional language extending $\mathbf{L_1}$, of which the additional atoms are the ground atoms defined by the following signature:

a finite set of attribute names:	$\mathbf{a_1, ... , a_m}$
a finite set of values for each of the attributes:	$\mathbf{v_{1,1}, ..., v_{1,k1}, v_{2,1}, ..., v_{m,km}}$

A variable over attributes will be denoted by **A**, a variable over values will be denoted by **V**.

Predicates:

is_incompatible_with(P, A, V)
has_value(A, V)
is_indicative(P)

The basic idea is that certain (observed) properties may rule out certain values for certain attributes. A fact **is_incompatible_with(P, A, V)** means that if the observed object has property **P**, then the attribute **A** of the object can not have the value **V**. The predicate **has_value(A, V)** means that attribute **A** of the object has value **V**. The last predicate requires a bit more explanation. The basic assumption on the domain is that we may have (potentially) many observations, which can be contradictory. That is, two observed properties may both rule out values for one attribute, such that together they rule out *all* possible values of that attribute. This may happen for a number of reasons. It may be that our observations are fallible: sometimes we observe a property the object does not have. It is also possible that our knowledge about which properties are incompatible with which values of attributes is uncertain or even not completely correct. Another possibility is that the object is not strictly delineated or strictly homogeneous with respect to its attributes, and some properties are observed from different parts of the object. To deal with this situation, we may label some observed properties as being *indicative*. If the observations are uncertain, 'indicative' may simply mean 'assumed true'. If the object is not homogeneous, then an indicative property is a property related to the view on the object we are interested in. The idea is that some properties are used to infer the values of attributes (in this sense they are 'indicative' of these values), whereas the others are for example wrong, not of interest or coincidental for this view.

There is a knowledge base, **KB**, in language $\mathbf{L_2}$, that consists of propositional formulae expressing knowledge which is of the following form:

- a (large) number of ground instances of:

  is_incompatible_with(P, A, V)

These instances represent the experts' knowledge of which properties rule out which values of certain attribute values.

- all ground instances of the generic rule

  $$\textbf{is\_indicative(P)} \wedge \textbf{is\_incompatible\_with(P, A, V)} \rightarrow \neg\, \textbf{has\_value(A, V)}$$

This rule makes it possible to conclude that certain attributes of the object do not have a certain value. This derivation can be made if an indicative property has been found that does not (generally) occur in objects for which the attribute **A** has value **V**.

- statements expressing that for each attribute at least one value should apply

  $$\textbf{has\_value}(\mathbf{a_1}, \mathbf{v_{1,1}}) \vee \ldots \vee \textbf{has\_value}(\mathbf{a_1}, \mathbf{v_{1,k1}})$$

  ....

  $$\textbf{has\_value}(\mathbf{a_m}, \mathbf{v_{m,1}}) \vee \ldots \vee \textbf{has\_value}(\mathbf{a_m}, \mathbf{v_{m,km}})$$

For a given set of observed properties $\mathbf{OBS} \subseteq \mathbf{Props}$, i.e., input of the form

$$\{ \text{observed}(p) \mid p \in \text{OBS} \}$$

the set

$$X = KB \cup \{ \text{is\_indicative}(p) \mid p \in \text{OBS} \}$$

may be inconsistent. That is, it may be inconsistent to assume that all observed properties are indicative for the object. This may occur if there is an attribute **A** such that for all of its possible values $\mathbf{V_{jk}}$, a property **P** is observed that negatively indicates this value (which means we have both **is_indicative(P)** and **is_incompatible_with(P, A, $\mathbf{V_{jk}}$)**). With the generic rule, the conclusion **¬ has_value(A, $\mathbf{V_{j,k}}$)** is drawn for all possible values $\mathbf{V_{j,k}}$ of **A**. But this is inconsistent with the statement

$$\text{has\_value}(A, V_{j,,1}) \vee \ldots \vee \text{has\_value}(A, V_{j,,kj})$$

which is in **KB**. However, the set of maximal indicative subsets consistent with **KB** may be considered. This is defined as follows:

Definition 2.3 (Maximal indicative subset)

a) A set of properties $S \subseteq \mathbf{Props}$ is an *indicative set of properties* if the theory

$$KB \cup \{\text{is\_indicative}(p) \mid p \in S\}$$

is consistent.

b) Let $\mathbf{OBS} \subseteq \mathbf{Props}$ be a given set of observed properties. A set $\mathbf{S} \subseteq \mathbf{OBS}$ is a *maximal indicative subset of* **OBS** if it is an indicative set of properties and for each indicative set of properties **T** with $\mathbf{S} \subseteq \mathbf{T} \subseteq \mathbf{OBS}$ it holds $\mathbf{S} = \mathbf{T}$.

The *set of maximal indicative subsets* of **OBS** is denoted by **maxind(OBS)**.

Note that, since **Props** is finite, for each indicative subset **S** of a set **OBS**, there exists at least one maximal indicate subset **S'** of **OBS** such that $\mathbf{S} \subseteq \mathbf{S'}$. Moreover, if **OBS** is an indicative set of properties itself, there is only one maximal indicative subset of **OBS**, namely **OBS** itself.

Based on these notions the following multi-interpretation operator can be defined.

Definition 2.4 (Generic multi-interpretation operator for approximate classification)

For a set $\mathbf{X} \subseteq \mathbf{L_1}$, define the *set of observations implied by* **X** by

$$OBS(X) = \{p \mid \text{observed}(p) \in Cn(X)\}.$$

The *operator* $\mathbf{MI_{maxind}}$ is defined by

$$MI_{maxind}(X) = \{ Cn(X \cup KB \cup \{\text{is\_indicative}(p) \mid p \in S\}) \mid S \in \text{maxind}(OBS(X)) \}$$

for each $\mathbf{X} \subseteq \mathbf{L_1}$.

Note that $\mathbf{X} \subseteq \mathbf{Y} \subseteq \mathbf{L_1}$ implies $\mathbf{OBS(X)} \subseteq \mathbf{OBS(Y)}$ Actually, the sets **X** will often be sets of the form **{observed(p) | p ∈ OBS}** for some set of properties $\mathbf{OBS} \subseteq \mathbf{Props}$.

2.3 Properties of the generic multi-interpretation operator for approximate classification

The operator $\mathbf{MI_{maxind}}$ satisfies a number of properties of well-behavedness. The proofs are rather straightforward.

Proposition 2.5

The multi-interpretation operator $\mathbf{MI_{maxind}}$ satisfies inclusion and non-inclusiveness.

In [5], some further conditions of well-behavedness for belief set operators are introduced (generalizing corresponding properties of inference operations). These properties can be defined for multi-interpretation operators as well; a number of them are formulated below.

Definition 2.6 (Properties of multi-interpretation operators)

a) Let $\mathcal{A}, \mathcal{B}$ be sets of belief sets. The set $\mathcal{B}$ *contains more information than* $\mathcal{A}$, denoted $\mathcal{A} \leq \mathcal{B}$, if for all $T \in \mathcal{B}$ there exists $\mathbf{S} \in \mathcal{A}$ such that $\mathbf{S} \subseteq \mathbf{T}$.

b) Let **MI** be a multi-interpretation operator.

MI satisfies *belief monotony* if for all $\mathbf{X, Y \subseteq L_1}$:

$$\mathbf{X \subseteq Y \;\Rightarrow\; MI(X) \leq MI(Y)}$$

c) Let **MI** be a multi-interpretation operator for which $\mathbf{L_1 \subseteq L_2}$.

1. **MI** satisfies *weak belief monotony* if for all $\mathbf{X, Y \subseteq L_1}$:

$$\mathbf{X \subseteq Y \subseteq K_{MI}(X) \Rightarrow MI(X) \leq MI(Y)}$$

2. **MI** satisfies *belief transitivity* if for all $\mathbf{X, Y, T \subseteq L_1}$:

$$\mathbf{T \in MI(X)\ \&\ X \subseteq Y \subseteq T \Rightarrow K_{MI}(Y) \subseteq T}$$

3. **MI** satisfies *belief cut* if for all $\mathbf{X, Y \subseteq L_1}$:

$$\mathbf{X \subseteq Y \subseteq K_{MI}(X) \Rightarrow MI(Y) \leq MI(X)}$$

Apart from belief monotony (which should in general not be expected), our multi-interpretation operator is well-behaved.

Theorem 2.7

The multi-interpretation operator $\mathbf{MI_{maxind}}$ satisfies weak belief monotony, belief transitivity and belief cut. It does not generally satisfy belief monotony.

Proof

Abbreviate $\mathbf{MI_{maxind}}$ to **MI**. Starting with belief monotony, consider a situation in which we have two properties, $\mathbf{P_1}$ and $\mathbf{P_2}$ (for simplicity), and suppose **KB** contains information which prevents $\mathbf{P_1}$ and $\mathbf{P_2}$ of both being indicative at the same time: there is an attribute **A** which has possible values 0 and 1. This means that **KB** contains the formula **has_value(A, 0)** $\vee$ **has_value(A, 1).** Furthermore, suppose that we have **is_incompatible_with**($\mathbf{P_1}$**, A, 0)** and **is_incompatible_with**($\mathbf{P_2}$**, A, 1)** in **KB**. Now let

$\mathbf{X = \{ observed(P_1) \}}$,

$\mathbf{Y = \{ observed(P_1), observed(P_2) \}}$.

Then **MI(X)** contains one element (in which $\mathbf{P_1}$ is indicative), and **MI(Y)** contains two elements, one in which only $\mathbf{P_1}$ is indicative, and one in which only $\mathbf{P_2}$ is indicative. For this latter element there is no smaller set in **MI(X)**. Therefore, belief monotony does not hold.

Let us now consider weak belief monotony and belief cut. Suppose $\mathbf{X \subseteq Y \subseteq K_{MI}(X)}$ and let $\mathbf{T \in MI(X)}$, then

$$\mathbf{T = Cn(X \cup KB \cup \{is\_indicative(p) \mid p \in M\})}$$

for some **M** $\in$ **maxind(OBS(X))** and **Y** $\subseteq$ **T** (since **Y** $\subseteq$ $\mathbf{K_{MI}(X)}$). But as **X** and **Y** contain only the predicate **observed** which is not present in **KB** or in **{is_indicative(p) | p** $\in$ **M}**, it must be the case that **Cn(Y)** $\subseteq$ **Cn(X)**, so that **Cn(X) = Cn(Y)**. This implies that **MI(X) = MI(Y)**, proving both weak belief monotony and belief cut.
If **T** $\in$ **MI(X) & X** $\subseteq$ **Y** $\subseteq$ **T**, then the same argument shows that **MI(X) = MI(Y)**, from which immediately follows that $\mathbf{K_{MI}(Y) = K_{MI}(X)} \subseteq \mathbf{T}$. This proves belief transitivity. ■

Each of the belief sets is an approximation in the sense of an upper bound of the solution. If the number of observations increases, this upper bound decreases, as is established in the following theorem.

Theorem 2.8
For each pair of subsets **X, Y** $\subseteq$ $\mathbf{L_1}$ the following holds:

$$\mathbf{X} \subseteq \mathbf{Y} \Rightarrow \text{ for all } \mathbf{S} \in \mathbf{MI(X)} \text{ there exists a } \mathbf{T} \in \mathbf{MI(Y)} \text{ such that } \mathbf{S} \subseteq \mathbf{T}$$

Proof
From **X** $\subseteq$ **Y** it follows **OBS(X)** $\subseteq$ **OBS(Y)** (see note just below Definition 2.4) Therefore every maximal indicative subset of **OBS(X)** is an indicative subset **S** of **OBS(Y)**. Within **OBS(Y)** this indicative subset can be extended to a maximal indicative subset **S'** (see note just below Definition 2.3). This implies the theorem. ■

This theorem guarantees that an increasing sequence of observations

$$\mathbf{X_0 \subseteq X_1 \subseteq X_2 \subseteq ....}$$

results in increasing beliefs sets within the sets $\mathbf{MI(X_i)}$**.** These increasing belief sets correspond to decreasing sets of classifications, i.e., for each of the increasing belief sets the ranges of the possible values of attributes are decreasing: this provides an approximation of the classification by a sequence of decreasing upper bounds.

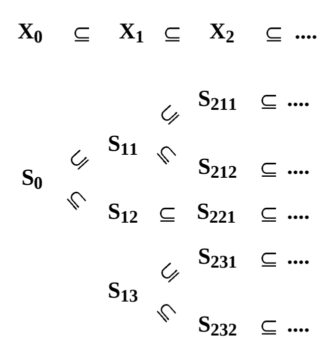

Fig. 1. Example approximate classificationsbased on an increasing sequence of observations

Note that Theorem 2.8 leaves open the possibility that belief sets remain constant, or new belief sets arise in some stage, i.e., sets of which no sub-set occurs in the previous set of belief sets. In general, for a given sequence of observations the resulting belief sets will form a set of trees as depicted in Figure 1. Here

$$\mathbf{MI(X_0) = \{S_0\}}$$
$$\mathbf{MI(X_1) = \{S_{11}, S_{12}, S_{13}\}}$$
$$\mathbf{MI(X_2) = \{S_{211}, S_{212}, S_{221}, S_{231}, S_{232}\}}$$

The following proposition covers the case of an observed set of properties **OBS** which has a unique interpretation:

Proposition 2.9

For each subset of properties $\mathbf{OBS} \subseteq \mathbf{Props}$ the following are equivalent:

(i) $\mathbf{MI_{maxind}}$ **({observed(p) | p** $\in$ **OBS})** contains just one element.

(ii) the set **OBS** is an indicative set of properties.

If these (equivalent) conditions are satisfied, all observed properties are indicative, and there are no alternative interpretations. This means there is no need for further selection from alternatives. The possible values of the attributes are contained in $\mathbf{MI_{maxind}}$ **({observed(p) | p** $\in$ **OBS})**.

If $\mathbf{MI_{maxind}}$ **({observed(p) | p** $\in$ **OBS})** contains more than one element, then a further selection process can be started. But even before this selection process, conclusions can be drawn: the kernel of the $\mathbf{MI_{maxind}}$ operator contains the most certain conclusions, so $\mathbf{K_{MI_{maxind}}}$ **({observed(p) | p** $\in$ **OBS})** may be inspected. For instance, there may be two possible views in $\mathbf{MI_{maxind}}$ **({observed(p) | p** $\in$ **OBS})** due to the fact that there is an attribute $\mathbf{A_1}$ for which no value is compatible with all the observed properties. However, all of these properties may indicate that another attribute $\mathbf{A_2}$ must have a certain value, and this conclusion will be in $\mathbf{K_{MI_{maxind}}}$ **({observed(p) | p** $\in$ **OBS})**. If $\mathbf{A_2}$ is all one is interested in, there is no need for selection. If one is interested also in $\mathbf{A_1}$, this selection has to take place. If one is interested in the properties which are indicative in both maximal indicative sets, one can either examine $\mathbf{K_{MI_{maxind}}}$ **({observed(p) | p** $\in$ **OBS})**, or the intersection of the maximal indicative sets:

$$\mathbf{K_{MI_{maxind}}(X) \cap \{\, is\_indicative(p) \mid p \in P \,\} = \{\, is\_indicative(p) \mid p \in \bigcap maxind(OBS(X)) \,\}}.$$

For the multi-interpretation operator $\mathbf{MI_{maxind}}$, the language and the format (the kinds of rules) of the knowledge base **KB** were fixed. When the language and format of **KB** is left open, we get a general class of multi-interpretation operators that can deal with input which is contradictory in the sense that it is inconsistent with a knowledge base.

3 An Example Application

In this section we will briefly describe a domain to which the formalization above was applied (see [1]). Nature conservationists are interested in a number of so-called *abiotic factors* of terrains. These factors, examples of which are the moisture, acidity and nutrient value, give an indication of how healthy a terrain is. As these factors are difficult to measure directly, a sample of plant species growing on a terrain is taken.

	Moisture						Acidity					Nutrient Value			
Species	vd	fd	fm	vm	fw	vw	bas	neu	sac	fac	ac	np	fnr	nr	vnr
Angelica sylvestris				x	x		x	x					x	x	
Carex acutiformis				x	x		x	x					x	x	
Carex riparia				x	x	x	x	x						x	x
Cirsium oleraceum				x	x		x	x					x	x	
Phalaris arundinacea			x	x	x	x	x	x						x	x
Phleum pratense ssp pratense			x	x			x	x						x	x
Poa trivialis			x	x	x		x	x						x	x
Caltha palustris ssp palustris				x	x		x	x	x			x	x	x	
Carex acuta				x	x	x	x	x	x				x	x	x
Cirsium palustre				x			x	x	x			x	x	x	
Crepis paludosa			x	x	x		x	x	x				x	x	
Deschampsia caespitosa			x	x	x		x	x	x				x	x	x
Epilobium parviflorum			x	x			x	x	x				x	x	
Equisetum palustre			x	x	x	x	x	x	x			x	x	x	
Filipendula ulmaria				x			x	x	x			x	x	x	
Galium palustre				x	x		x	x	x			x	x	x	x
Glyceria fluitans				x	x	x	x	x	x	x			x	x	x
Juncus articulatus				x	x		x	x	x			x	x	x	x
Lathyrus pratensis			x	x			x	x	x				x	x	
Lotus uliginosus			x	x	x		x	x	x			x	x	x	
Lychnis flos cuculi				x	x		x	x	x				x	x	
Lysimachia vulgaris			x	x	x		x	x	x			x	x	x	
Myosotis palustris				x	x		x	x	x				x	x	x
Scirpus sylvaticus				x	x	x	x	x	x				x	x	
Anthoxanthum odoratum		x	x	x					x	x		x	x		
Carex nigra			x	x	x				x	x	x	x	x		
Carex panicea			x	x	x				x	x		x	x		
Epilobium palustre			x	x	x				x			x	x		
Juncus conglomeratus		x	x	x					x	x		x	x		

Moisture	(vd: very dry, fd: fairly dry, fm: fairly moist, vm: very moist, fw: fairly wet, vw: very wet)
Acidity	(bas: basis, neu: neutral, sac: slightly acid, fac: fairly acid, ac: acid)
Nutrient value	(np: nutrient poor, fnr: fairly nutrient rich, nr: nutrient rich, vnr: very nutrient rich)

Table 1. Maximal indicative subsets within an inhomogeneous sample of plant species.

For each species, the experts have knowledge about the possible values of the abiotic factors of a terrain on which the species lives. So it may be known, for example, that a certain species can only live on medium to very acid terrains. Combining such knowledge for each of the plant species observed on a terrain leads to conclusions about the abiotic factors of the terrain.

During the development of a knowledge-based system, EKS, to automate this classification process, however, it turned out that the samples of species taken were often incompatible (e.g., see the sample depicted in Table 1). That is, there was at least one abiotic factor for which no value could be found that was permissible for all species. This is not due to errors in the knowledge of abiotic factors needed by species to live, but due to other effects. For example, a terrain may lie on the transition of a dry and a wet piece of land. Some of the observed species may occur on the drier, and others on the wetter side. This can also be due to the presence of ponds in an otherwise dry terrain. Also transitions of a terrain over time, or vertical inhomogeneity may be causes.

The approximate classification task described above is an example of the task performed by the multi-interpretation operator $\mathbf{MI_{maxind}}$. The object to be studied is a terrain, and the attributes of interest are the abiotic factors. The presence of certain species are the observable properties of the object. So we can specialize the generic knowledge base to this case. The language is as follows:

properties: (occurrence of) plant species names	**achillea_millefolium,**
	achillea_ptarmica, ...
attributes: abiotic factors	**moisture, acidity, nutrient_value**
values for each of the attributes abiotic factors:	**very_dry, fairly_dry, ...,**
	basic, neutral, ...,
	nutrient_poor, fairly_nutrient_rich, ...

The experts' knowledge about the possible values of abiotic factors for a species, is formalized by (a large number of) instances of the predicate **is_incompatible_with(P, A, V)**, where **P** is one of the plant species names, **A** is an abiotic factor, and **V** is a value of that factor. This knowledge leads to a specific instantiation of $\mathbf{MI_{maxind}}$, we will denote by the same name. The alternative interpretations given by $\mathbf{MI_{maxind}(X)}$ are extremely useful. Each of the alternatives leads to a different set of (possible) values for the abiotic factors. If this is for instance due to the fact that the terrain consists of a drier portion and a wetter portion, then a selection can be made for the portion of interest, whose possible values for abiotic factors are contained in the corresponding interpretation. This selection process can be formalized by a selection operator as defined in Definition 2.2. At this moment, that process has not been analyzed in more detail, but that is one of the future directions of research.

As mentioned before, a system called EKS has been developed to help a user in establishing the abiotic factors of a terrain. The correspondence between the formalization of the expert reasoning task and the interactive knowledge-based system EKS that models the approximate classification task is as follows (see Figure 2).

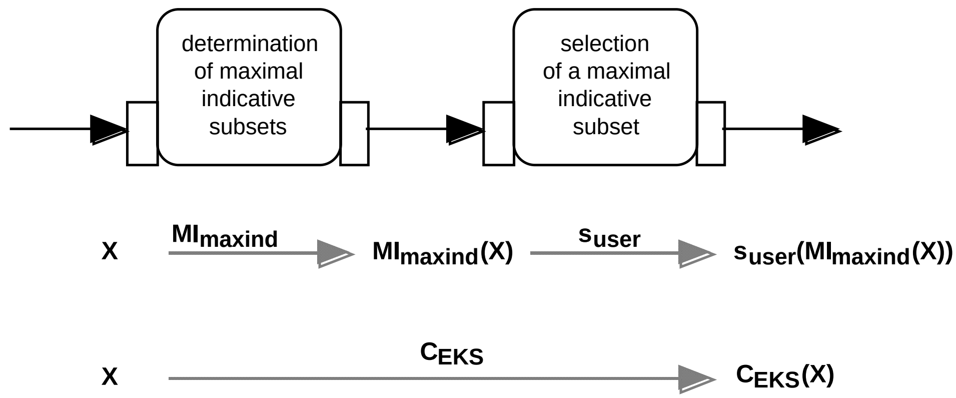

Figure 2. Correspondence between the formalization and the system.

The first component of the system, determination of maximal indicative subsets, is formalized by the belief set operator $\mathbf{MI_{maxind}}$ defined in Section 2.2. The second component of the system, selection of a maximal indicative subset, which models (an interface to) the selection process by the user of the system, is formalized by a single-valued selection function $\mathbf{s_{user}}$.

The composition $\mathbf{C_{EKS}}$ of $\mathbf{MI_{maxind}}$ and $\mathbf{s_{user}}$ defined by

$$\mathbf{C_{EKS}(X) = s_{user}\,(MI_{maxind}\,(X))} \qquad \text{for } \mathbf{X \subseteq L_1}$$

is a selective interpretation operator for $\mathbf{MI_{maxind}}$ (as described in Definition 2.2b). This operator formalizes the reasoning of the system in interaction with the user as a whole. Note that from the two functions of which this overall function is composed, one is fixed and defined by the system itself (i.e., $\mathbf{MI_{maxind}}$), whereas the other can be changed dynamically, depending on the user (i.e., $\mathbf{s_{user}}$). For more details on this application, see [1].

4 Representation in Default Logic

The previous section described the generic multi-interpretation operator **MI**, which formalizes the interpretation of (possibly inconsistent) observation information using maximal indicative sets. A specification of this multi-interpretation operator in a (well-known) logical formalism would mean that known results about this logic can be applied to this situation, but it would also allow for the use of proof mechanisms for this logic to be used in an implemented system based on such an operator. In [6] and [9] default logic is used as a specification language for families of belief sets. These results can be applied to the formalization of the previous section.

To start, a brief overview of Reiter's default logic (cf. [2], [11]) is provided. Although Reiter's definitions are stated for any first-order language, here they are restricted to propositional logic, as is commonly done. So let us again assume a propositional language **L**. A *default rule* (or *default*) is an expression of the form $(\alpha : \beta_1, ..., \beta_n) / \gamma$ where $\alpha, \beta_1, ..., \beta_n$ and γ are propositional formulae. Intuitively such a default rule means: if α is believed and it is not inconsistent to assume β_1 through β_n, then assume γ. A *default theory* Δ is then a pair $< \mathbf{W}, \mathbf{D} >$ with **W** a set of sentences (the *axioms* of Δ) and **D** a set of default rules. The default rules are used to extend the axioms to a (larger) set of formulas, called an *extension*. The following definition of the notion of extension is slightly different but equivalent to Reiter's original definition.

Definition 4.1 (Reiter extension)

Let $\Delta = < \mathbf{W}, \mathbf{D} >$ be a default theory. A set of sentences **E** is called a *Reiter extension* of Δ if the following condition is satisfied:

$$\mathbf{E} = \bigcup_{i=0}^{\infty} \mathbf{E_i}$$

where

$$\mathbf{E_0 = Cn(W)},$$

and for all $\mathbf{i \geq 0}$

$$\mathbf{E_{i+1} = Cn(E_i \cup \{ \omega \mid (\alpha : \beta_1,...,\beta_n) / \omega \in D, \alpha \in E_i \text{ and } \neg \beta_1 \notin E, ... , \neg \beta_n \notin E \})}$$

The *set of Reiter extensions* of Δ is denoted by $\mathbf{Ext(\Delta)}$.

Extensions of a default theory are closed under propositional provability, so $\mathbf{Ext(\Delta)}$ is a family of belief sets. In a sense, this family is represented (or specified) by Δ. For an arbitrary family of belief sets, the question can be posed whether it can be represented by a default theory.

Definition 4.2 (Representability of a family of belief sets)

Let $\Delta = < \mathbf{W}, \mathbf{D} >$ be a default theory. A family of belief sets **F** is *representable by* Δ if $\mathbf{Ext(\Delta) = F}$. The family **F** is called *representable by a default theory* if there exists such a default theory.

In [9] the following theorem has been proven (Corollary 5.2):

Theorem 4.3

A family **F** of theories is representable by a normal default theory if and only if $\mathbf{F = \{L\}}$ or there is a consistent set of formulas **W** and a set of formulas **C** such that

$$\mathbf{F = \{ Cn(W \cup \Phi) \mid \Phi \text{ is a maximal subset of } C \text{ consistent with } W \}}$$

In [6] the question is posed whether a belief set operator can be represented by a set of defaults. Below, the definitions in that paper are slightly generalized to deal with a different input and output language. Recall that $\mathbf{L_1}$ is the input language, and $\mathbf{L_2}$ is the output language. We make the assumption that $\mathbf{L_1 \subseteq L_2}$.

Definition 4.4 (Representability of a multi-interpretation operator)

Let $\Delta = < \mathbf{W}, \mathbf{D} >$ be a default theory. A multi-interpretation operator **MI** is *representable by* Δ, if for all $\mathbf{X} \subseteq \mathbf{L_1}$ it holds that $\mathbf{MI(X) = Ext(< W \cup X, D >)}$. The operator **MI** is called *representable by a default theory* if there exists such a default theory.

Consider the family of belief sets $\mathbf{MI_{maxind}(X)}$ where $\mathbf{X} \subseteq \mathbf{L_1}$. Then Theorem 3.3 can be applied to $\mathbf{MI_{maxind}(X)}$ by setting:

$$\mathbf{W = X \cup KB}$$
$$\mathbf{C = \{ is\_indicative(p) \mid p \in OBS(X)\}}$$

whith **KB** as defined in Section 2.2. Now Theorem 4.3 implies that for each $\mathbf{X} \subseteq \mathbf{L_1}$ there exists a normal default theory that represents the belief sets of $\mathbf{MI_{maxind}(X)}$. The theorem does not imply that there exists *one* set of defaults **D** which works for all sets $\mathbf{X} \subseteq \mathbf{L_1}$, so this does not imply that the multi-interpretation operator $\mathbf{MI_{maxind}}$ is representable by a default theory. However, the normal default theory can actually be found by defining the following generic set of defaults **D**:

(observed(p) : is_indicative(p)) / is_indicative(p) for all properties **p** in **Props**.

This set of defaults is independent of **X**, so $\mathbf{MI_{maxind}}$ is representable using the above set of defaults **D** and the **KB** of Section 2.2.

Theorem 4.5

The multi-interpretation operator $\mathbf{MI_{maxind}}$ is representable by the normal default theory **< KB, D >**.

Proof

Let **X** be a set of formulas in $\mathbf{L_1}$. Let $\mathbf{X \cup KB}$ be consistent (if it is not, verification is straightforward and omitted). The extensions of $< \mathbf{KB \cup X, D} >$ are sets of the form $\mathbf{Cn(KB \cup X \cup S)}$, where **S** is a subset of $\mathbf{\{ is\_indicative(p) \mid observed(p) \in Cn(X) \}}$, which is maximal such that $\mathbf{Cn(KB \cup X \cup S)}$ is consistent. This is proved below. The sets $\mathbf{Cn(KB \cup X \cup S)}$ with **S** as above together comprise $\mathbf{MI_{maxind}(X)}$.
First of all, let **S** be such a maximal set, and let $\mathbf{E = Cn(KB \cup X \cup S)}$. Then if the $\mathbf{E_i}$ are defined as in Definition 4.1, the following holds:

$$\mathbf{E_0 = Cn(KB \cup X),}$$
$$\mathbf{E_1 = Cn(E_0 \cup \{ is\_indicative(p) \mid observed(p) \in E_0 , \neg\, is\_indicative(p) \notin E \})}$$

As $\mathbf{E_1}$ does not contain more instances of the **observed** predicate than $\mathbf{E_0}$ (this follows from the fact that **X** contains only the **observed** predicate, whereas **KB** does not), $\mathbf{E_i = E_1}$ for all $\mathbf{i > 1}$. The claim is that

$$\mathbf{\{ is\_indicative(p) \mid observed(p) \in E_0 , \neg\, is\_indicative(p) \notin E \} = S.}$$

Suppose **observed(p)** $\in$ $\mathbf{E_0}$ and $\neg$ **is_indicative(p)** $\notin$ **E.** Then **observed(p)** is in **Cn(X)** and **Cn(KB** $\cup$ **X** $\cup$ **S** $\cup$ **{ is_indicative(p) })** is consistent. But as **S** was maximal with respect to these properties, **is_indicative(p)** $\in$ **S**. On the other hand, if **is_indicative(p)** $\in$ **S**, then **observed(p)** $\in$ $\mathbf{E_0}$ and $\neg$ **is_indicative(p)** $\notin$ **E** (as **E = Cn(KB** $\cup$ **X** $\cup$ **S)** is consistent).
Now let **E** be an extension of **< KB** $\cup$ **X, D >**, then it is of the form **Cn(KB** $\cup$ **X** $\cup$ **S)**, where **S** contains (only) formulas of the form **is_indicative(p)**. Examination of **KB** (and the restriction on the language of **X**), shows that only if **observed(p)** $\in$ **Cn(X)** is **is_indicative(p)** $\in$ **E**. As extensions are always consistent (if each rule has a justification and the axioms are consistent), **Cn(KB** $\cup$ **X** $\cup$ **S)** must be consistent. Suppose there exists a **T** $\supset$ **S** (strict inclusion) respecting the conditions, then there must be a default rule **observed(p) : is_indicative(p) / is_indicative(p)**, with **observed(p)** $\in$ **Cn(X)** $\subseteq$ **E** and **Cn(KB** $\cup$ **X** $\cup$ **S** $\cup$ **{ is_indicative(p) })** consistent, implying that $\neg$ **is_indicative(p)** $\notin$ **E**. But that means there is an applicable default rule for which the conclusion is not in **E**, contradicting the assumption that **E** is an extension. Therefore **S** must be maximal. ■

At this point the reader may wonder what the benefit is of the representation in default logic. The multi-interpretation operator $\mathbf{MI_{maxind}}$ arose during the analysis and formalization of the application described in the previous section. The system, EKS, was designed and implemented based on this operator $\mathbf{MI_{maxind}}$. The implementation in fact follows the definition (Definition 2.4) rather closely. The results of the current section indicate that alternatively a theorem prover for default logic (or, rather, a program computing extensions of default theories) could be used. A highly optimised theorem prover for default logic obviates the need to optimise this part of the system ourselves. This is the subject of current work on the system.

5 Discussion

In most real-life classification problems, the information about the object to be classified can be interpreted in different ways. In this paper, multi-interpretation operators were introduced to formalize this interpretation process. In particular, observation results of the world may underspecify or overspecify a classification. Overspecification means that the observations are in contradiction with knowledge about the world. A generic multi-interpretation operator was introduced for approximate classification tasks where attribute values of an object are determined on the basis of imperfect interpretation of observable properties of the object. The multi-interpretation operator formalizes in a neat manner the different variants of approximate classifications of the object. This operator is rather well-behaved, and can be represented by a default theory. This can be a basis for the use of (highly optimized) theorem provers for default logic, to implement a system formalized by the multi-interpretation operator. For the domain of ecological classification an application for the theory has been developed, and the resulting system, EKS, that has been implemented has shown to be a useful tool for nature conservationists.

After multiple interpretations of observation information have been identified, often a choice is made for one of them. Which view is (or which views are) most appropriate presumably requires additional heuristic (strategic) knowledge (cf. [3], [4], [12]). One of the areas of future research is to further analyze this choice process, in general terms, but also in particular for the knowledge-based system. Future research will focus on the acquisition of this knowledge to be able to support users in the selection process.

Acknowledgements

Within the EKS-project (funded by IPTS and Staatsbosbeheer) a number of persons have contributed their expertise: Frank Cornelissen, Edgar Vonk (scientific programmers); Christine Bel, Rineke Verbrugge (for parts of the analysis of the domain); Bert Hennipman and Frits van Beusekom (domain experts from IPTS); Piet Schipper, and Wim Zeeman (domain experts from Staatsbosbeheer).

References

[1] F.M.T. Brazier, J. Engelfriet, and J. Treur, "Analysis of Multi-interpretable Ecological Monitoring Information", In: A. Hunter, S. Parsons (eds.), *Applications of Uncertainty Formalisms*, Lecture Notes in AI, vol. 1455, Springer Verlag, 1998, pp. 303-324. Extended version in: *Applied Artificial Intelligence Journal*, vol. 16, 2002, pp. 51-71.

[2] P. Besnard, *An Introduction to Default Logic*, Springer-Verlag, 1989.

[3] G. Brewka, "Adding Priorities and Specificity to Default Logic", in: C. MacNish, D. Pearce, L.M. Pereira (eds.), *Logics in Artificial Intelligence, Proceedings of the JELIA-94*, Lecture Notes in Artificial Intelligence, vol. 838, Springer-Verlag, 1994, pp. 247-260.

[4] G. Brewka, "Reasoning about Priorities in Default Logic", in: *Proceedings of the AAAI-94*, 1994.

[5] J. Engelfriet, H. Herre and J. Treur, "Nonmonotonic Reasoning with Multiple Belief Sets", *Annals of Mathematics and Artificial Intelligence*, vol. 24, 1998, pp. 225-248. Preliminary version in: D.M. Gabbay, H.J. Ohlbach (eds.), *Practical Reasoning, Proceedings FAPR'96*, Lecture Notes in Artificial Intelligence, vol. 1085, Springer-Verlag, 1996, pp. 331-344.

[6] J. Engelfriet, V.W. Marek, J. Treur and M. Truszczynski, Default Logic and Specification of Nonmonotonic Reasoning. *Journal of Experimental and Theoretical AI*, vol. 13, 2001, pp. 99-112. Preliminary version in: J.J. Alferes, L.M. Pereira, E. Orlowska (eds.), *Logics in Artificial Intelligence, Proceedings of the Fourth European Workshop on Logics in AI, JELIA'96*, Lecture Notes in Artificial Intelligence, vol. 1126, Springer-Verlag, 1996, pp. 224-236.

[7] W. Hodges, *Model theory*, Cambridge University Press, 1993.

[8] D. Makinson, "General Patterns in Nonmonotonic Reasoning", in: D.M. Gabbay, C.J. Hogger, J.A. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3*, Oxford Science Publications, 1994, pp. 35-110.

[9] V.W. Marek, J. Treur and M. Truszczynski, "Representation Theory for Default Logic", *Annals of Mathematics and Artificial Intelligence* 21, 1997, pp. 343-358.

[10] V.W. Marek and M. Truszczynski, *Nonmonotonic logics; context-dependent reasoning*, Springer-Verlag, 1993.

[11] R. Reiter, "A Logic for Default Reasoning", *Artificial Intelligence* 13, 1980, pp. 81-132.

[12] Y.-H. Tan, J. Treur, "Constructive Default Logic and the Control of Defeasible Reasoning", in: B. Neumann (ed.), *Proceedings of the European Conference on Artificial Intelligence, ECAI'92*, John Wiley and Sons, 1992, pp. 299-303.