An Inductive Logic Programming Framework to Learn a Concept from Ambiguous Examples

Dominique BOUTHINON[1,2] Henry SOLDANO[1,2]
[1] Atelier de BioInformatique (ABI)
[2] Laboratoire d'Informatique Paris Nord (LIPN)

Atelier de BioInformatique 12, rue Cuvier 75005 PARIS (FRANCE)
email : Dominique.Bouthinon@snv.jussieu.fr
phone : (33) 01 44 24 65 82
fax : (33) 01 44 27 63 12

Abstract. We address a learning problem with the following peculiarity : we search for characteristic features common to a learning set of objects related to a target concept. In particular we approach the cases where descriptions of objects are ambiguous : they represent several incompatible realities. Ambiguity arises because each description only contains indirect information from which assumptions can be derived about the object. We suppose here that a set of constraints allows the identification of "coherent" sub-descriptions inside each object.
We formally study this problem, using an Inductive Logic Programming framework close to characteristic induction from interpretations. In particular, we exhibit conditions which allow a pruned search of the space of concepts. Additionally we propose a method in which a set of hypothetical examples is explicitly calculated for each object prior to learning. The method is used with promising results to search for secondary substructures common to a set of RNA sequences.

1. Introduction

A problem of molecular biology is in the origin of this study, namely the one of finding a secondary substructure common to a set of RNA sequences[2,3]. A major feature of this problem is that, as often, learning consists in searching for a characteristic description related to the target-concept, but only an ambiguous representation of each object is known. More precisely when considering the whole description of an object there are several, and possibly a great number, of incompatible realities. This occurs either because various realities exist (in Dietrich [6] the object is a molecule which has various, alternative, 3-D conformations), either because, as the observation is indirect, a unique reality is dissimulated among a set of possibilities. The RNA problem mentioned above corresponds to this later case. But let us take a simpler example : the object is a scene figuring a stoplight, the description, due to a color-blind person is the following: "*Stoplight(s), Red(s), Green(s), Car(v), Run(v,s)*". The difficulty is that the observation is imperfect (because the witness is color-blind) and indirect (as run(s) is an interpretation rather than a direct perception). There are two incompatible realities depending whether the stoplight were either Red or Green. In the former case the object should be described as "*Stoplight(s), Red(s),Car(v), run(v,s)*", in the latter case "*Stoplight(s), Green(s), Car(v)*". Suppose now that this color-blind person builds up a set of such ambiguous observations related to a target concept (for instance the responsibility of a driver in a car crash), each observation containing a single positive instance. The learning problem would be to find a characteristic description, belonging to the hypothesis space, i.e. a most specific hypothesis that explains all positive instances. Note that in the ordinary case an incoherent

statement as "Red(s), Green(s)" could be generated but would be rejected since it could not recognize any positive instance. It is not the case when considering ambiguous observations. Then, how to avoid such hypothesis? What we propose here is to use integrity constraints, as for instance " *Stoplight(s) and Red(s) and Green(s),* is incoherent ". We state that part of an observation is "coherent" if it satisfies all the constraints, and that an hypothesis is coherent if it does not recognize any incoherent *sub-observation* (a part of any observation).
The constraints allows the identification, within an observation, of the largest coherent sub-observations, i.e. the possible realities on which induction should be performed. We have here the inductive view of the idea of « extension » as used in non-monotonic reasoning [5,17]. For now on we will refer to a learning problem with ambiguous examples and integrity constraints as *a class A problem.*
The framework described in section 2 relies on logic programming, with a closed-world assumption (within a sub-observation anything which is not stated nor can be deduced is false) together with negation by failure when expressing constraints. In section 3 we show that such ambiguities within examples result in several difficulties about completeness and consistency when searching a space of concepts and studying properties allowing for an efficient reduction of the search space. Then we propose a general method to perform concept learning in which we first build and select, the set of largest coherent sub-observations for each example. Section 4 addresses the RNA secondary structure problem mentioned at the beginning of the paper. We describe the application of our method to this problem and discuss some encouraging results. We conclude by discussing related work, in particular the work on the «multiple instance problem » as described by Dietterich [6,10].

2. Class A problems

A class A problem formally consists in learning from a set O of observations, and, given a domain theory T, a set H of characteristic representations of a target concept. Each observation **o** contains one example **e**(o) of the target concept. Our framework is embedded in Inductive Logic Programming [7,11,12], and from now on, H, O, and T will be logic programs without functional terms. We allow the use of a negation by failure operator, **not**, within the body of clauses belonging to the domain theory T.

An observation **o** is represented as a conjunction of ground atoms: $\mathbf{o} = \mathbf{a}_1 \wedge \mathbf{a}_2 \wedge \mathbf{a}_m$. The example **e**(o) hidden inside **o** is a particular sub-observation, i.e. a part of **o**. The definitions of the concepts in H are searched inside a space of hypothesis. Any hypothesis **h** of this space is represented as a conjunction of single clauses (ground or not) ($\mathbf{h} = \mathbf{b}_1 \wedge \mathbf{b}_2 \wedge ...\mathbf{b}_n$).
The domain theory T contains a set C of constraints together with a set B of definite clauses describing background knowledge about constraints. Each constraint is a goal, i.e. a negative clause $\leftarrow \mathbf{d1} \wedge \mathbf{d2} \wedge ...\mathbf{d_u}$ where $\mathbf{d}_i$ is a (possibly negative) literal. The idea here is that $\mathbf{d1} \wedge \mathbf{d2} \wedge ...\mathbf{d_u}$ represents an incoherent statement :

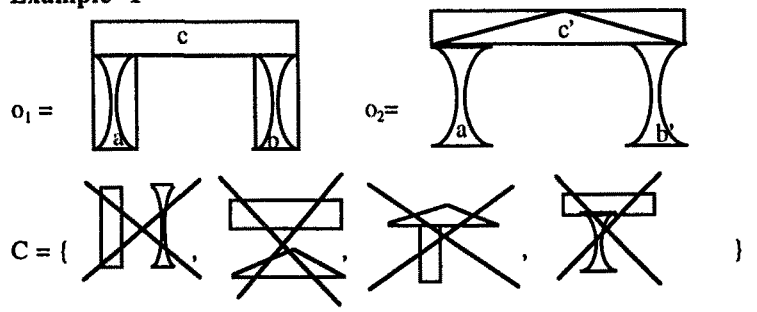
Definition 1 *a sub-observation o is incoherent iff o violates a constraint* $c = « \leftarrow d1 \wedge d2 \wedge ...d_u »$ *i.e. iff the goal* $\leftarrow d1 \wedge d2 \wedge ...d_u$ *succeeds within the logical program* $o \wedge B$.

Generally observations are incoherent and examples, by definition, are not.

As defined the constraints in a class A problem are close to integrity constraints as presented by De Raedt [14] and used in various works [18-21,26].

Hereafter we exemplify our framework using an « arch problem » similar to the famous problem due to Winston [25]. Arches are hidden inside the observations, as could happen if each arch were represented by a picture suffering from unexpected superposition

Example 1



The domain theory T contains :

- the following constraints:

c_1 = ← *column(X)* ∧ *pillar(Y)* (there cannot be columns and pillar in the same arch)
c_2 = ← lintel(X) ∧ triangle(Y) (there cannot be lintels and triangles in the same arch)
c_3 = ← **incompatibility** (this constraints is defined thanks to the background knowledge)

- The background knowledge

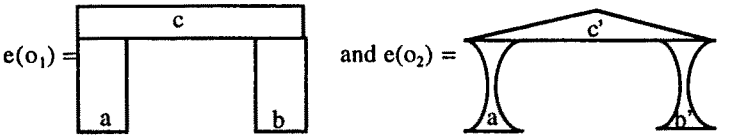
B = {incompatibility ← pillar(X) ∧ triangle(Y) ∧ sustains(X,Y),
incompatibility ← column(X) ∧ lintel(Y) ∧ sustains(X,Y)}
(there is an incompatibility when a pillar sustains a triangle, or when a column sustains a lintel)

$\{o_1,o_2\}$ is the set of observations :

o_1 = ***column(a)*** ∧ *pillar(a)* ∧ *column(b)* ∧ *pillar(b)* ∧ **lintel(c)** ∧ **sustains(a,c)** ∧ sustains(b,c).
o_2 = **column(a')** ∧ **column(b')** ∧ **lintel(c')** ∧ triangle(c') ∧ **sustains(a',c')** ∧ sustains(b',c').

Within the observations the conjunction of italic atoms and the conjunction of bold atoms represent incoherent sub-observations (not all ones) which violate the constraints printed with the corresponding typographic style. For instance o_1 violates c_1 and c_3 since there are pillars and columns within o_1, i.e. the goal =← *column(X)* ∧ *pillar(Y)* succeeds within the logic program o_1 ∧ B with the answer-substitutions {X/a, X/b}, and there is an incompatibility due to the fact that a column (a or b) sustains a lintel (c), i.e. the goal <- incompatibility succeeds within o_1 ∧ B.

The observations o_1 and o_2 each contains an example corresponding to an (unidentified) coherent sub-observation:



End example 1

A class A problem is then represented by the pair (O, T = B ∪ C). The purpose is then to build *the set H of most specific hypothesis that "cover" all the examples* (completeness).

We state that an hypothesis **h** is complete relatively to the set O of observations, if and only if **h** theta-subsume [4,9,13] each example within O. In the same way, the hypothesis **h** is *more general* than the hypothesis **g** (denoted as **h** ≤ **g**) whenever **h** theta-subsumes **g**.

Definition 2 *An hypothesis **h** is complete* $\Leftrightarrow \forall o \in O, \exists \theta$ *such that* $h\theta \subseteq e(o)$ *(denoted as* $h \leq e(o)$*).*
Here, for sake of clarity, the domain theory is not used to check completeness. However there would be no conceptual difficulties in using generalized subsumption as proposed by Buntime [4].

However as examples are hidden within observations, completeness checking is difficult : an hypothesis **h** can cover an observation and still not cover the embedded example. Constraints are then used to ensure that a part of such absurd (relatively to the domain theory) hypothesis are not selected. Hereafter we define the incoherence of an hypothesis relative to the set of constraints :

Definition 3 *An hypothesis **h** is incoherent iff there exists a substitution* λ *such that* $h\lambda$ *corresponds to some incoherent sub-observation* ($\exists\lambda\ \exists o \in O$, $h\lambda \subseteq o$ *and* $h\lambda$ *incoherent*) .

Example 2
let us go back to example 1. The sub-observation o'= ***column(b)*** ∧ **lintel(c)**∧ **sustains(b,c)** included in o_1 is incoherent. As a consequence the hypothesis *h* = ***column(X)*** ∧ **lintel(Y)**∧ **sustains(X,Y)** is incoherent because when λ = {X/b,Y/c}, we have o' =hλ.
End example 2

The set H we search is then the set of most specific hypothesis both complete and coherent. Notice that, whenever there is no ambiguity, i.e. the examples are exactly the observations, then observations have to be coherent. However incoherence still could eliminate hypothesis that can be unified with an incoherent part of some example (the whole example being coherent). As we will see an interesting case (hereafter referred to as "monotonicity") is the one in which whenever a sub-observation is coherent no part of it can be incoherent.

3. Solving a class A problem

As stated above direct use of completeness ($\forall o \in O$, $\mathbf{h} \leq \mathbf{e}(o)$) is in this case prohibited since e(o) is unknown. So, we cannot build the set H of the most specific complete and coherent hypothesis we search for. Our first purpose is then to determine a part G of the hypothesis space, as small as possible, which necessarily includes H ($H \subseteq G$).
In what follows we present such a set G together with some conditions under which an efficient pruning of the search space is possible.

3.1 Bounding the set of solutions

We determine hereafter a set L of hypothesis which is a lower bound of H. Later we will discuss to what extent searching for L is interesting when solving class A problems. The first step consists in weakening completeness in order to have a computable criteria.

Definition **4** An hypothesis **g** is stated as O-complete if and only if **g** theta-subsumes each observation ($\forall \mathbf{o} \in O$, $\mathbf{g} \leq \mathbf{o}$).

A O-complete and coherent hypothesis is said to be O-*valid*, or, in short, *valid*. The set G of valid hypothesis includes H since any complete hypothesis necessarily is O-complete ($\forall \mathbf{o} \in \mathbf{O}$, $\mathbf{h} \leq \mathbf{e}(o) \Rightarrow \mathbf{h} \leq \mathbf{o}$).
Let us denote as L the set of the most specific valid hypothesis. Let **h** be an hypothesis belonging to H, then there exists at least one hypothesis in L which is more specific than **h** (consequently no hypothesis of L can be strictly more general than **h**). So L is a lower bound of H within the set of valid hypothesis. Of course when the examples are exactly the observations, then there is no ambiguity and so L = H. However an interesting case is the one in which we assume that H is included in L. We will give some rationale for such an assumption in section 3.3 and argue that this assumption holds when considering the real-world problem presented in 4.
Hereafter we consider top-down methods to search for G. As G is bounded by L these methods also apply when only searching for L. The starting point is the (always valid) null hypothesis .

3.2 Pruning the top-down search

An O-complete but incoherent hypothesis happen to be specialized in a valid hypothesis 1) by grounding variables within **h**, 2) by adding atoms to **h**. Generally speaking **h** is specialized in $\mathbf{h'} = \mathbf{h}\pi \wedge \mathbf{b}$ where π is a substitution and **b** a conjunction of atoms.

Example 3
Let us consider the following class A problem (O, T= B $\cup$ C) :
O ={o} with **o = true(a,3)** $\wedge$ **implies(a,b)** $\wedge$ **true**(b,5) $\wedge$ *true(c,7).*
C = {$\leftarrow$ *true(c,X),* $\leftarrow$ **true(X,U)** $\wedge$ **implies(X,Y)** $\wedge$ **not true(Y,V)**}.
B = $\varnothing$.
1) The hypothesis **h** = *true(X,U)* is incoherent as **h** can be unified to the incoherent sub-observation *true(c,7)*. However the hypothesis **h'** = true(a,U), more specific than **h** (by grounding X) is coherent.
2) The hypothesis **g = true(X,U)** $\wedge$ **implies(X,Y)** is incoherent as **g** and the sub-observation o' = **true(a,3)** $\wedge$ **implies(a,b)** are unifiable (o' is incoherent as the goal $\leftarrow$ **true(X,U)** $\wedge$ **implies(X,Y)** $\wedge$ **not true(Y,V)** succeeds in (**o'** $\wedge$ B)). Here again the

hypothesis **g'** = true(X,U) ∧ implies(X,Y) ∧ true(Y,V), more specific than **g** (by adding true(Y,V)), is coherent.
End example 3

The situation shown in the example 3 prohibits any pruning relying on validity. We study hereafter the cases in which such situations cannot occur and so pruning becomes possible. For that purpose we use the framework of Torre and Rouveirol [23, 24] who states that a property P is « private », relatively to a relation R whenever for any pair $(\mathbf{h_1},\mathbf{h_2})$, $\{\mathbf{h_1}$ R $\mathbf{h_2}$ and $\neg P(\mathbf{h_1})\}$ implies $\neg P(\mathbf{h_2})$. In our case, asserting that validity is private relatively to the subsumption relation, means that any hypothesis that is not valid will not have valid specializations. We exhibit hereafter two conditions which are sufficient to insure that validity is private relatively to subsumption. Notice that example-1/2 satisfies both condition when example-3 satisfies none of them.

The first condition, *'incoherence discrimination'*, is denoted as D(O,T).

***Definition** 5 D(O,T) is true whenever for any observation, two sub-observations which are identical except for constant names, are either both coherent or both incoherent.*

The second condition *'incoherence monotony'* is denoted as M(O,T) :

***Definition** 6 M(O,T) is true whenever any sub-observation including an incoherent sub-observation also is incoherent.*

Considering example 3, D(O,T) does not hold since **o** contains the sub-observation *true(a,3)* which is coherent and also contains *true(c,7)* which is incoherent. M(O,T) neither holds since the sub-observation **true(a,3)** ∧ **implies(a,b)** ∧ true(b,5) is coherent and includes the incoherent sub-observation **true(a,3)** ∧ **implies(a,b)**. Here falsity of M(O,T) is related to negative literals appearing in constraints.
So D(O,T) appears to limit the ambiguity tolerated within an observation when M(O,T) restricts the expressiveness of constraints.
The following property states that D(O,T) and M(O,T) are sufficient to insure that validity is private relatively to subsumption (proofs of properties are given in appendix) :

***Property** 1 If M(O,T) and D(O,T) both hold then whenever an hypothesis **h** is not valid, none of its specializations **h'** are valid.*

Now suppose that M(O,T) and D(O,T) both hold. Then, when considering top-down search of the hypothesis space, specialization of an hypothesis can be safely stopped whenever it is not O-complete or it is incoherent.

3.3 Prior construction of hypothetical examples

M(O,T) together with D(O,T) suggest a top-down search starting from the null hypothesis (always valid) and maintaining at each step a set of valid hypothesis.
A natural idea when dealing for a class A problem is to associate to each observation **o** the set E(o) *of hypothetical examples* i.e. the largest coherent sub-observations of **o**. One can observe that each hypothetical example represents a coherent and not ambiguous part of the

data associated with a given observation. In our framework, using hypothetical examples has several advantages. First, a given incoherent sub-observation causes incoherence of many hypothesis (by way of substitutions) and so prior computation of largest coherent sub-observations avoid unnecessary calculations. Second, some class A problems are naturally expressed as « multiple instances » problems, as we will see in section 4. Third, reformulating ambiguity in terms of alternatives results in a different view of class A problems. Note that as examples are supposed to be coherent, e(o) is included in at last one hypothetical example.
First we give two properties that relate incoherence monotony and incoherence discrimination to hypothetical examples.

Property 2 *If M(O,T) holds then a sub-observation o' of a given observation o is coherent if and only if it is included in at least one hypothetical example of o.*

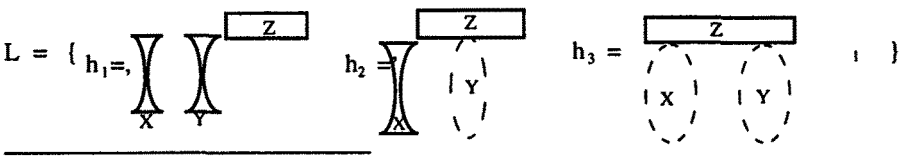
Property 3 *if M(O,T) and D(O,T) both hold then an hypothesis h is valid if and only if h theta subsumes at least one hypothetical example of each observation o*

Now a general way to solve a class A problems requires two steps:
1) Build the set E(o) corresponding to each observation o. This is performed by classical branch and bound methods whose bound part is based on the integrity constraints.
2) Build the set G of valid hypothesis. This is performed by classical Top-Down methods using property 3.

During step 2 it is possible, whenever we assume that H is included in the lower bound L of G, to only retain hypothesis belonging to L. This means that no hypothesis more specific than any hypothesis h belonging to H (h is a most specific complete and coherent hypothesis) can subsume at least one hypothetical example of each observation. Whenever possible, sorting G (or L) using a problem dependent preference criterion helps in selecting most « plausible » hypotheses.

Example 4

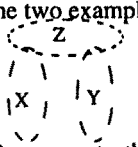
M(O,T) and D(O,T) both hold when considering example-1. We also suppose that we have at our disposal the two sets $E(o_1)$ and $E(o_2)$ of hypothetical examples respectively related to o_1 and o_2 Furthermore a top-down search starting from the null hypothesis and using $E(o_1)$ and $E(o_2)$ has selected a set G of valid hypothesis, the following set L containing the most specific ones ($L \subseteq G$) :



[1] the dashed line means that the shape of considered element the arch is not known.

(h_1= column(X) ∧ column(Y) ∧ lintel(Z)
h_2= column(X) ∧ lintel(Z)∧ sustains(Y,Z)
h_3= lintel(Z) ∧ sustains(X,Z) ∧ sustains(Y,Z))

In this case, the solution set H only contains one most specific hypothesis h that covers the two examples e(o1) and e(o2) (h = sustains(X,Z) ∧ sustains(Y,Z) representing



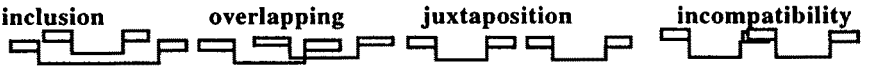
)

On can note that h is valid and complete but does not belong to the set L of the most specific valid and O-complete hypothesis. This shows that an important final ambiguity remains since H must be searched in G. Thus a plausibility criterion must be applied on G to select one (of few) preferred hypothesis.

End example 4

4. Application

The theoretical study we have presented is inspired in a biomolecular problem concerning the prediction of the secondary structure common to a set of RNA molecules. A RNA is a molecule made of a *sequence* of nucleotides, that is folded in such a way that some nucleotides are paired (two paired nucleotides constitute a *base-pairing*), in several places named *helices*. The helices of a RNA determinate its *secondary structure* which is determinant to the biological function of the molecule. It is very hard, often impossible, to experimentally identify RNAs secondary structures. Therefore, one has to compute all potential helices of each RNA sequence using theoretical methods. These potential helices constitute an indirect and ambiguous view of the secondary structure of a RNA: indeed some helices are incompatible among each other, thus the actual secondary structure is one of the (numerous) sets of compatibles helices. Hence, predicting the secondary structure common to a set of RNA molecules is a class A problem: for each RNA the set of all potential helices is an observation, the hypothetical examples are the largest sets of compatible helices, and the example is one particular set of compatible helices. The domain theory of the problem only contains the single compatibility constraint between helices.
Two helices can have one of the following four structural relations (an helix is represented by the following scheme )

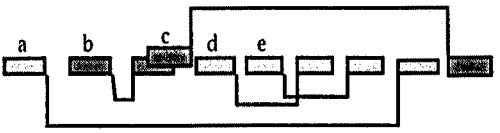


We are only interested by the structural shape of the secondary structures, so an observation describing all the structural relationships between potential helices of a RNA will contain a conjunction of atoms written from the four binary predicates « juxtaposition », « overlapping », « inclusion », « incompatibility ». The single constraint is represented by the goal ← *incompatible(X,Y)*.

Therefore any *completely defined sub-observation* (i.e. a sub-observation describing all the relationships between helices it mentions) that does not contain the atom « incompatible » represents a potential secondary structure since the goal ← *incompatible(X,Y)* fails in this context.

Example 5

Let the 5 potential helixes of a RNA sequence s, with the following structural relations (one can observe that the helices **c** and **d** are incompatible) be:



The hypothetical examples contain, respectively, all the relations of the helixes {a,b,c,e} and {a,b,d,e}.Let the observation describing the relationships between the helices of s be :
inclusion(a,b) ∧ overlapping(a,c) ∧ inclusion(a,d) ∧ inclusion(a,e) ∧
incompatible(b,c) ∧ juxtaposition(b,d) ∧ juxtaposition(b,e) ∧
inclusion(c,d) ∧ inclusion(c,e) ∧ overlapping(d,e).
Let the constraint be ← *incompatible(X,Y)*, and the knowledge associated with the domain be empty (B = ∅), then each sub-observation containing the atom *incompatible* is incoherent. One observes that the sub-observation *inclusion(a,b)* ∧ *inclusion(a,c)* which is formally coherent is not completely defined due to the absence of the relationship « incompatible(b,c) » (the sub-observation inclusion(a,b) ∧ inclusion(a,c) ∧ incompatible(b,c) is completely defined and incoherent).

End example 5

In this problem, we will only consider the completely defined sub-observations reducing the space of the hypothesis to the ones unifiable with such sub-observations (all the properties described in the theoretical part remaining true).
The solution set H contains only one hypothesis **h** representing the actual secondary substructure common to all RNAs. Moreover we will suppose that H is included in the set L containing the most specific valid hypothesis (any valid hypothesis represents a structure that is potentially present in all RNAs), which means that **h** is one of the largest (most specific) potential common substructures.
This problem satisfies the conditions D(O,T) and M(O,T), thus we have designed a new method of prediction using the scheme of resolution presented in part 3. We expose hereafter the main ideas of this method (cf. [2,3] for the details) :
In a preliminary step we build the observations from the potential helices (of each sequence) identified by an ad hoc program [1]. At the first step we compute the hypothetical examples E(o) of each observation **o**, that represent the secondary structures potentially present in the RNA associated with **o**, through a branch and bound algorithm using the constraint of compatibility. Because of the cardinality of E(o) we only keep the 200 potential structures with best energies in each observation (remark that this criterion cannot be use to directly search for the common secondary structure).
At the second step, a top-down ad hoc method is used to build the set containing the largest potential secondary structures common to all RNAs, i.e. the sets L of the valid

hypothesis that are maximally specific. We have shown that thanks to a new way of encoding secondary structures, building L is equivalent to searching for the longest prefixes in a dictionary, where each entry of the dictionary represents a potential structure. This is done with an algorithm derived from KMR [8,22], that we have adapted.
At the third step we sort L with a plausibility measure that consists in comparing for each candidate structure represented by a single hypothesis **h** of L, the number of RNA sequences covered by **h** with the number of previously built random sequences also covered by **h**.
This new method presents interesting features because it does not require the prior alignment of the sequences[2] and allows taking into account all theoretical shapes of secondary structures. The first results are promising : we applied the method to several sets of sequences (about 200 in a set), each sequence containing one transfer RNA (tRNA) and non relevant data. We were able to discover the tRNA structure in all the sets.

5. Conclusion

As stated above a major feature of class A problems is the ambiguity of examples which are considered as hidden in observations. More precisely the part of the observation outside each example is not only useless (not related to a target concept) but does not correspond to the reality.
As a consequence inside an observation we may find incoherent parts resulting either from uncertainties as for instance "heavy(x) and light(x)" or from misunderstandings as « dead(x) and smiling(x) ». Then background knowledge has to include constraints to prevent considering such incoherent parts. We have seen above that ambiguity together with incoherence modify the statement of the classical "learning a characteristic concept description" problem, and that, when certain properties are satisfied, a convenient strategy consists in first building the sets of hypothetical examples related to each observation.
Class A problems are close to the «multiple instance problem» as described by Diettrich, but differ from it in two main points. First, in class A problems we have to compute hypothetical examples, Diettrich assumes direct access to « multiple » instances. Diettrich assumes that there are various realities : in this problem a molecule has several conformations, some of which are related to the target-concept, and these conformations are not mixed up, but correspond to different descriptions. Second, Diettrich uses counter-examples to constraint the search of discriminant features, in our framework this search is more difficult because we do not have such knowledge.

We are now working on extending the class A framework with negative observations (each containing a negative example), and generalized subsumption. Moreover we are studying links between our framework and *characteristic induction from interpretations* [15,16].

Acknowledgments

We thanks Alain Viari and Marc Champesme for their contribution to this work, and Eduardo Rocha for his help in carefully reading this manuscript.

Appendix

Proof of property 1 :

[2]Aligning two RNA sequences is equivalent to establish a map between the sets containing the positions of their nucleotides.

h incoherent $\Rightarrow \exists o \subseteq O$, $\exists \sigma$ such that $h\sigma \subseteq o$ and $h\sigma$ incoherent. Let us suppose that **h'** is valid (consequently O-complete and coherent). $h \leq h' \Rightarrow h' = h\pi \wedge b$, where π is a substitution and **b** a conjunction of *atoms*. **h'** O-complete $\Rightarrow \exists \theta$ such that $(h\pi \wedge b)\theta \subseteq o \Rightarrow h\pi\theta \subseteq o$. D(O,T) and $h\sigma \subseteq o$ and $h\pi\theta \subseteq o$ and $h\sigma$ incoherent $\Rightarrow h\pi\theta$ incoherent (Definition 5). $h\pi\theta$ incoherent and $h\pi\theta \subseteq (h\pi \wedge b)\theta \subseteq o \Rightarrow (h\pi \wedge b)\theta$ incoherent (Definition 6), thus **h'** is incoherent (consequently not valid) ◊

Proof of property 2 :
"$\Rightarrow$" : **o'** being coherent either **o'** is maximal (thus $o' \in E(o)$), either $\exists e \in E(o)$ such that $o' \subset e$.
"$\Leftarrow$" : Let $o' \subseteq e$. Because **e** is coherent (by definition) the property M(O,T) entails that **o'** is coherent (otherwise **e** should not) ◊

Proof of property 3 :
"$\Rightarrow$" : **h** O-complete entails $\forall o \in O$, $h \leq o \Rightarrow \exists \theta$ such that $h\theta \subseteq o$. As **h** is coherent the property 2 entails that $\exists e \in E(o)$ such that $h\theta \subseteq e$, so $h \leq e$.
"$\Leftarrow$" : Let $\forall o \in O\ \exists e \in E(o)$ such that $h \leq e$ be. We want to proof that a) $h \leq o$ and b) **h** is coherent :
a) $\exists e \in E(o)$ such that $h \leq e \Rightarrow \exists \theta\ h\theta \subseteq e$, as $e \subseteq o$ we can derive $h\theta \subseteq e \subseteq o$, what entails $h \leq o$.
b) if **h** is incoherent then $\exists o \in O$ and $\exists \theta'$ such that $h\theta' \subseteq o$ and $h\theta'$ is incoherent. According to a) $\exists \theta$ such that $h\theta \subseteq e \subseteq o$, and $h\theta$ is coherent (property 2). Thus $h\theta'$ is incoherent and $h\theta$ is coherent what contradicts the property D(O, T), thus $h\theta'$ (consequently **h**) is not incoherent ◊

References

1. B. Billoud, M. Kontic and A. Viari, Palingol: a declarative programming language to describe nucleic acids' secondary structures and to scan sequence databases, *Nucleic Acids Research*, vol 24, n° 8, pp. 1395-1403, 1996.
2. D. Bouthinon, Apprentissage à Partir d'Exemples Ambigus : Etude Théorique and Application à la Découverte de Structures Communes à un Ensemble de Séquences d'ARN, doctorat de l'université Paris XIII-Institut Galilée, 1996.
3. D. Bouthinon, H. Soldano and B. Billoud, Apprentissage d'un concept commun à un ensemble d'objets dont la description est hypothétique : application à la découverte de structures secondaires d'ARN, *Journées Francophones d'Apprentissage*, Sètes (France), pp. 206-220, 1996.
4. W. Buntime, Generalized subsumption and its applications to induction and redundancy, *Artificial Intelligence*, pp. 149-176, 1988.
5. D. MacDermott and J. Doyle, Non-monotonic logic I, *Artificial Intelligence*, vol. 13, n° 1-2, pp.41-72, 1980.
6. T.G. Dietterich, R.H. Lathrop and T. Lozano-Perez, Solving the Multiple-Instance Problem with Axis-Parallel Rectangles, *Artificial Intelligence*, vol 89(1-2), pp. 31-71, 1997.
7. P.A. Flach, A Framework for Inductive Logic Programming, *Inductive Logic Programming*, 1992.

8. R.M. Karp, R.E. Miller and A.L. Rosenberg, Rapid identification of repeated patterns in strings, trees and arrays, *Proc. 4th Annu. ACM Symp. Theory of Computing*, pp. 125-136, 1972.
9. J.U. Kietz and M. Lübbe, An efficient subsumption algorithm for inductive logic programming, *Machine Learning*, pp. 130-138, 1994.
10. R. Lindsay, B. Buchanan, E. Feigenbaum and J. Lederberg, Applications of artificial intelligence to organic chemistry : The Dendra projects, New-York McGraw-Hill., 1980.
11. S. Muggleton, Inductive Logic Programming, *Inductive Logic Programming*, 1992.
12. S. Muggleton and C. Feng, Efficient induction of logic programs, *Inductive Logic Programming*, 1992.
13. S.H. Nienhuys-Cheng and R.D. Wolf, The subsumption theorem in inductive logic programming : facts and fallacies, *International Workshop on Inductive Logic Programming*, Leuven, 1995.
14. L. De Raedt and M. Bruynooghe, Belief updating from integrity constraints and queries, *Artificial Intelligence*, vol 53, pp. 291-307, 1992.
15. L. De Raedt, Logical settings for concept-learning, *Artificial Intelligence*, 95 187-201 (1997).
16. L. De Raedt and L. Dehaspe, Clausal Discovery, *Machine Learning*, 26, 99-146 (1997).
17. R. Reiter, A logic for default reasoning, *Artificial Intelligence*, vol. 13, n° 1-2, pp. 81-131, 1980.
18. M. Sebag, A constraint-based induction algorithm in FOL, *Machine Learning*, pp. 275-283, 1994.
19. M. Sebag and C. Rouveirol, Induction of maximally general clauses consistent with integrity constraints, *Inductive logic programming*, 1994.
20. M. Sebag and C. Rouveirol, Constraint inductive logic programming, *Advances in inductive logic programming*, 1995.
21. M. Sebag, C. Rouveirol and J.F. Puget, Inductive Logic Programming + Stochastic Bias = Polynomial Approximate Learning, Multi Strategy Learning, J. Wrete & R.S. Michalski, MIT Press, 1996.
22. H. Soldano, A. Viari and M. Champesme, Searching for flexible repeated patterns using a non transitive similarity relation, *Pattern Recognition Letters*, vol 16, pp. 233-245, 1995.
23. F. Torre and C. Rouveirol, Opérateurs naturels en programmation logique inductive, JICAA' 97, Roscoff 20-22 Mai 1997, pp. 53-64, 1997.
24. F. Torre and C. Rouveirol, Natural ideal operators in inductive logic programming, *Proceeding of the ninth European Conference on Machine Learning* pp. 274-289, Springer-Verlag, 1997.
25. P.H. Winston, Learning structural descriptions from examples, *The psychology of computer vision*, P.H. Winston, MacGraw-Hill, 1975.
26. S. Wrobel, First order theory refinement, Advances in inductive logic programming, L. De Raedt, IOS press, 1996.