INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS 1
Int. J. Intel. Systems XX, yyy-zzz (1999)

CONSTRUCTIVE REINFORCEMENT LEARNING

JOSE HERNANDEZ-ORALLO*

Department of Information Systems and Computation, Technical University of Valencia, Camí de Vera 14, Aptat. 22.012 E-46071, Valencia, Spain

ABSTRACT

This paper presents an operative measure of reinforcement for constructive learning methods, i.e., eager learning methods using highly expressible (or universal) representation languages. These evaluation tools allow a further insight in the study of the growth of knowledge, theory revision and abduction. The final approach is based on an apportionment of credit wrt. the 'course' that the evidence makes through the learnt theory. Our measure of reinforcement is shown to be justified by cross-validation and by the connection with other successful evaluation criteria, like the MDL principle. Finally, the relation with the classical view of reinforcement is studied, where the actions of an intelligent system can be rewarded or penalised, and we discuss whether this should affect our distribution of reinforcement. The most important result of this paper is that the way we distribute reinforcement into knowledge results in a rated ontology, instead of a single prior distribution. Therefore, this detailed information can be exploited for guiding the space search of inductive learning algorithms. Likewise, knowledge revision may be done to the part of the theory which is not justified by the evidence. © XXXX John Wiley & Sons, Ltd.

KEY WORDS: Reinforcement Learning; Theory Evaluation; Incremental Learning; Ontology; Apportionment of Credit; Abduction; Induction; MDL principle; Knowledge Acquisition and Revision; ILP; Philosophy of Science

1. INTRODUCTION

1.1. Motivation

The problem of inductive learning defined as "the construction of theories that describe the evidence" is *underspecified*. Consequently, many evaluation criteria have been presented to complete this specification. Model 'simplicity' and 'reinforcement' are the most natural ones and they have been successfully applied to restricted representations. However, they do not scale up well to constructive languages, i.e., languages that allow dynamical change of its representational bias (what is sometimes known as the possibility of 'redescription'[21]). The issue is especially troublesome when new constructed terms can be created to express more compactly the evidence (this is usually known in ILP as the problem of predicate invention).

Some approximations have been adopted to make the simplicity criterion work for constructive (or universal) languages (e.g. U-learnability[35]). Contrariwise, a constructive and general formalisation for measuring reinforcement has not been presented to date, despite the fact that reinforcement learning in restricted representations (like general Markov decision processes[23]) has been especially

* Correspondence to: J. Hernandez-Orallo, Departament de Sistemes Informàtics i Computació, Universitat Politècnica de València, Camí de Vera 14, Aptat. 22.012 E-46071, València, Spain. E-mail: jorallo@dsic.upv.es. On-line papers: http://www.dsic.upv.es/ ~jorallo/ escrits/ escritsa.htm

© 1999 John Wiley & Sons, Ltd

fruitful in this decade (see e.g. References 20 and 43 for surveys) and it has been recently related with EBL[8].

The reasons may be found in the increasing difficulty of assigning and 'propagating' the reinforcement (or apportionment of credit[19]) depending on two factors: (1) the eagerness of the inductive strategy and (2) the expressibility of the language which is to be used for the hypotheses.

Since the expressibility of the representation language does not imply a constructive learning method (this expressive power cannot be exploited), for the rest of the paper we will define constructive learning as having these two characteristics: eager strategy and highly expressible representation languages.

Not surprisingly, these two issues are as well related. Eager learning methods extract all the regularity from the data in order to work with intensional knowledge (instead of the extensional knowledge of lazy methods[1]) but, on the other hand, intensional knowledge is only possible if the representation language is rich enough. The difficulty of these two issues explains the broad use of lazy methods, like instance-based and case-based reasoning[30], and algorithms for restricted representations, like attribute learning. In decision trees or attribute languages, no invented terms are induced and reinforcement is distributed among the initial attributes. The main drawback of these approaches is the lack of flexibility: when arrived at a 'saturation' point, the data are not abstracted further and the mean reinforcement cannot be increased. Furthermore, the ontology must be given and not constructed (a model of the 'world' is embedded in the system) and the possible extensions of this world are very restricted.

Inductive Logic Programming (ILP) is the best example of a recent reaction towards 'constructive learning', but other frameworks like Explanatory Based Learning (EBL) have tried to adopt reinforcement as an evaluation criterion[8]. The combination faces many difficulties apart from efficiency: a main problem is presented when learning in highly expressible frameworks: the ontology of any new constructed concept is indirect. The usual solution to this problem is the assumption of a prior probability. Once the probabilities are assigned, a bayesian framework can be used to 'propagate' the distribution.

In general, there is not any justification at all of which prior distribution to choose. In the absence of any knowledge, the most usual one is the MDL (Minimum Description Length) principle[37,38]. The MDL principle is just a formalisation of Occam's razor (the preference of the shortest theories). Theoretically, its close relation with PAC-learning[45] has been established by Blumer et al.[4] and, recently, some high-level representation inductive methods (e.g. U-learnability in ILP[35]) have adapted these ideas.

All of them are based on the assumption of a prior probability. However, there are many riddles associated with the management of probabilities and, in particular, the best choice, the MDL principle, has additional ones.

1.2. Proposal

As we will see, most of these difficulties would disappear if no prior distribution is assumed and the knowledge is constructed by reinforcement, as the data suggest. However, as we have discussed, the translation of these ideas to general representational frameworks seems difficult. First, the length of the structures which supposedly are to be reinforced is variable. Secondly, and more importantly, it seems we can always invent 'fantastic' concepts that can be used in the rest of knowledge. Consequently, these 'fantastic' concepts are highly reinforced, increasing the reinforcement ratio of knowledge in an unfair way.

© 1999 John Wiley & Sons, Ltd

An immediate way out is the combination of reinforcement learning with some prior probability, mainly the MDL principle (see e.g. an example under the name of 'incremental self-improvement'[39]), to restrict the appearance of these inventions. Notwithstanding, our approach also avoids 'fantastic' concepts but it is based exclusively on reinforcement. Consequently, compression turns out to be an 'a posteriori' consequence of a well-established reinforcement, instead of an 'arbitrary' assumption.

1.3. Paper Organisation

The paper is organised as follows. Section 2 presents some model selection criteria usually essayed in machine learning, especially the MDL principle. Section 3 introduces our framework for incremental knowledge construction. Section 4 essays a first adaptation of reinforcement to realise the problems of 'fantastic' concepts. Section 5 remakes the approach and introduces the idea of 'course' to measure reinforcement. Section 6 introduces some applications and examples. Section 7 gives a justification of reinforcement by relating cross-validation and intensionality. Section 8 considers the length of the reinforced 'units' or 'rules' and establishes the relation with the MDL principle in the limit. The result is a balanced reinforcement suitable for explanation, not so strict as the whole avoidance of exceptions (and noise). Section 9 discusses the extension of these ideas to wider notions of reinforcement with the presence of reward and penalties. Section 10 presents two methods for computing effectively these measures and deals with their limitations and complexity. Section 11 closes the paper discussing the results and the open questions.

2. SELECTION CRITERIA IN INDUCTIVE INFERENCE

The aim of Machine Learning is the computational construction of hypothetical inferences from facts, as Michalski has pointed out[31]: "*inductive inference was defined as a process of generating descriptions that imply original facts in the context of background knowledge. Such a general definition includes inductive generalisation and abduction as special cases*".

However, given the background knowledge B and some evidence E, infinite many hypotheses H can be induced which ensure $B \cup H \models E$. As we have said, some selection criterion is needed to complete the specification of the learning problem. Intrinsically, selection criteria can be classified in the following way:

- **Semantical Criteria** (*What does the hypothesis cover?*). There has been a long and still open debate among informativeness (advocated by Popper[36]), non-presumptiveness, generality, specificity, etc., apart from other considerations: the theory may be complete or partial, and exact or approximate.
- **Syntactical Criteria** (*What is the hypothesis like?*). Different ad-hoc preferences have been adopted depending on the purpose of the learning task. However, the MDL principle is the syntactical criterion that has been used more frequently. In general, syntactical criteria imply the assumption of a prior distribution which can be used to derive a likeliness value of hypotheses.
- **Structural Criteria** (*How does the hypothesis cover the evidence?*). The best known structural criterion is the computational complexity of a hypothesis (time). This criterion is implicitly assumed by the computational restrictions of the learning algorithm. Nonetheless there are other structural criteria which are much more interesting, around the ideas of Whewell's 'consilience'[48], Reichenbach's principle of common cause, Thagard's coherence[44], all of them vs. separate covering of the evidence. Unfortunately, there are not formalisations or even clear defini-

© 1999 John Wiley & Sons, Ltd

tions for these terms. The same applies for related concepts like intensionality vs. tolerance of extensionality (intrinsical exceptions).

In the rest of this section, we will center on the pros and cons of the principle of simplicity, because it is the one that has been better formalised. We will show later that it is closely related with reinforcement. The principle of simplicity, represented by Occam's razor, selects the shortest hypothesis as the most plausible one. This principle was rejected by Karl Popper because, in his opinion (and at that moment) there *was* no objective criterion for simplicity. However, Kolmogorov Complexity or Algorithmic Information[47], denoted $K(x)$, is an absolute criterion for simplicity. This is precisely what R.J.Solomonoff proposed as a 'perfect' theory of induction[29,42]. The direct relationship between Kolmogorov Complexity and Stochastic Complexity inspired J. Rissanen in 1978 to use it as a general modelling method, giving the popular MDL principle[37], recently revised as a one-part code[38] instead of the initial two-parts code formulation.

It is remarkable (and often forgotten) that Kolmogorov Complexity just gives consistency to this theory of induction; Occam's razor is *assumed* but not proven. Nonetheless, some justifications have been given in the context of physics, reliability and entropy, but, in our opinion, it is the notion of *reinforcement* (and cross-validation) which justifies the MDL principle more naturally. At first sight, it *seems* that the higher the mean compression ratio ($length(E)$ / $length(H)$) the higher the mean reinforcement ratio.

Summing up, the MDL principle says that, in absence of any other knowledge about the hypotheses distribution, we should select the prior $P(h) = 2^{-K(h)}$, prevailing short theories over large ones. However, this prior distribution has many riddles. First of all, (1) $K(h)$ is not computable, so it must be approximated (e.g. using the time-weighted variant Kt of Kolmogorov complexity[27]), with the additional problem that it may dynamically change as the learner knows that something can be further compressed. Second, (2) it presents problems with perfect data; the MDL principle usually 'underfits' the data, because sometimes it is too conservative for incremental learning. New examples are merely quoted until their compression is worthy. Third, (3) the reliability of the theory is not always increasing with the number of examples which have confirmed the theory (e.g., the sequence $(a^n b^n)^*$ is more compressible if $n = 10^{10}$ than if $n = 78450607356$). Moreover, (4) it is difficult to work with different and non-exclusive hypotheses, because, if we have T_a and T_b, intuition (and logic) says that $T = T_a \vee T_b$ should have more probability, but the MDL principle assigns less probability to T because it is larger. Finally, (5) the MDL principle has shown problems for explanation, because, for the sake of maximum mean compression, some part of the hypothesis cannot be compressed at all. This yields a very compressed part plus some additional extensional cases which are not validated, making the whole theory weak. An ontology is difficult to construct from here if they are unrelated (not explained) with the other facts. This is associated with the differentiation between Enumerative Induction and Best Explanation[14,15,9] and the distinction between Induction and Abduction[11,16].

We will handle these difficulties with a structural criterion: a dynamical and detailed propagation of reinforcement. Our approach has additional advantages: (1) no prior assumption has to be made (apart from how to distribute this reinforcement, which is the topic of this paper), i.e. knowledge is constructed just as the data suggest, and (2) reinforcement can be more flexibly managed than probabilities, and allows further insight on the relation between the evidence and the theory.

© 1999 John Wiley & Sons, Ltd

3 FRAMEWORK

With this section we just present the schema of incremental learning and the languages we aim to address in the following sections.

3.1. Incremental Knowledge Construction

From the complex task of knowledge construction, organisation and maintenance, for our purposes, we will exclusively tackle the inductive or learning task in knowledge construction.

Incremental knowledge construction (which includes acquisition and revision) generates a theory from an evidence that is gradually supplied example by example. From the very beginning, with an empty knowledge $T=\emptyset$, when new observations or evidences e are received, we can have three possible situations:

- **Prediction Hit** (*or 'matter of course'*). The observations are covered without more assumptions, i.e., $T \models e$. The theory T is reinforced.
- **Novelty**. The observation is uncovered but consistent with T, i.e. $T \not\models e$ and $T \cup e \not\models \square$. Here, the possible actions are:
  1. *Extension*: T can be extended with a good explanation A, (i.e. $T \cup A \models e$).
  2. *Revision*: revised if a good explanation cannot be found,
  3. *Patch*: quoted as an extensional exception (i.e. $T' = T \cup e$), or
  4. *Rejection*: regarded as noise and ignored.
- **Anomaly**. The observation is inconsistent with the theory T, i.e., $T \not\models e$ and $T \cup e \models \square$. In this case, T cannot be extended and there are three possibilities: *revision*, *patch* or *rejection*.

An eager but still non-explanatory approach to theory formation is Kuhn's theory of changing paradigms[26] which basically matches with the MDL principle: as too many exceptions to the paradigm are found, they are increasingly lengthy to quote (*patch*) and the whole paradigm (or part of it) must be reformulated.

In the preceding schema, abduction performs an extension of current knowledge with some assumption (usually one or more facts) and induction can be an extension or revision which performs some kind of generalisation. Nevertheless, this characterisation is not sufficient for a clear distinction (see Reference 16 for more detailed considerations). In fact, it is a topic of current discussion (for a state of the art see Reference 11). In this way, abduction has been commonly seen as belief revision[5], usually combined with induction[2]. In other cases it is related with validation, justification or ontology[10], in the way that the part of the theory where abduction is supported must be reliable. Unavoidably, this reliability must come from a reinforcement produced by the previous evidence.

The former schema is general enough to include explanatory and conservative knowledge construction. Explanatory knowledge construction should minimise the exceptions, so patches and rejections should not be allowed. Thus, revisions are much more frequent. Even more, the goal is to anticipate, to invest, to find more informative and easily refutable hypotheses[36], in contrast to what many approaches to minimal revisions aim for (see e.g. Reference 33), supported by the obvious fact that a minimal revision is usually less costly, *in short-term*, than a deep revision.

3.2. Representation Languages. Syntax and Semantics

For the study of reinforcement we need to introduce some basics for the representation to which it can be applied. A 'pattern' of languages is defined as a set of *chunks* or rules *r* which are composed of a head (or consequence) and a body (or set of conditions) in the following way $r \equiv \{ h :- t_1, t_2, .. t_s \}$. A theory is simply a set of rules: $T = \{r_1, r_2, \ldots, r_m\}$

© 1999 John Wiley & Sons, Ltd

Since no restriction of how h and t_i can be (there may be variables, equations, boolean operators...), our definition could be specialised to propositional languages, Horn theories, full logical theories, functional languages, some kind of grammars, and even higher-order languages. In the following, we leave unspecified the semantics of the representations and we just say that e is a consequence of P, denoted $P \models e$ (in other words, there is a proof for e in P, or, simply, P covers e).

3.3. Preliminaries

Given the slight semantical and syntactical restriction of the previous paragraphs, we introduce some useful and simple constructions which will shape our framework with more determination.

Definition 3.1. A rule r_i is said to be necessary wrt. T for an example e iff

$$T \models e \quad \text{and} \quad T - \{r_i\} \not\models e$$

Definition 3.2. A theory T is reduced for an example e iff

$$T \models e \quad \text{and} \quad \neg\exists\, r_i \in T \text{ such that } r_i \text{ is not necessary for } e$$

For the rest of the paper, we consider a proof as a set of rules, independently of their order of combination, the applied substitutions or the number of times that each rule is used. This unusual (and incomplete) conception of proof allows us to work without considering the concrete semantics while maintaining an appropriate degree of detail. This makes possible the following definition:

Definition 3.3. We say that S_1 and S_2 are alternative proofs for an example e in the theory T iff

$$S_1 \subset T,\ S_2 \subset T,\ S_1 \neq S_2 \text{ and } S_1 \text{ and } S_2 \text{ are reduced for } e$$

We denote with $Proof(e,T)$ the set of alternative proofs for an example e wrt. a theory T. Finally, we can define $Proof_r(e,T)$ as the set of alternative proofs which contain r. More formally,

Definition 3.4.

$$Proof_r(e,T) = \{\ S : S \subset Proof(e,T) \text{ and } r \in S\ \}$$

With these naive constructions, we are able to introduce our first measurement of reinforcement.

4. REINFORCEMENT WRT. THE THEORY USE

As we have seen, whatever the approach to knowledge construction, the revision of knowledge must come either from an inconsistency or from a lack of support. In the latter case, a partial or total weakness of the theory can be detected by a loss of *reinforcement* (or apportionment of credit[19]). There have been several empirical and theoretical justifications for reinforcement in different fields, from many empirical observations on learning processes in animals or humans to theoretical and practical verifications by cross-validation.

We present the first intuitive way to compute the reinforcement map for a given theory, depending on past observations.

Definition 4.1. The pure reinforcement $\rho\rho(r)$ of a rule r from a theory T wrt. some given evidence $E = \{e_1, e_2, \ldots, e_n\}$ is defined as:

$$\rho\rho(r) = \Sigma_{i=1..n}\, card(Proof_r(e_i,T))$$

In other words, $\rho\rho(r)$ is computed as the number of proofs of e_i where r is used. If there are more than one proof for a given e_i, *all* of them are reckoned, but in the same proof, a rule is computed only once.

Definition 4.2. The (normalised) reinforcement is defined as:

$$\rho(r) = 1 - 2^{-\rho\rho(r)}$$

Definition 4.2 is motivated by the convenience of maintaining reinforcement between 0 and 1. However, its computation is easy, as the following elementary lemma shows:

© 1999 John Wiley & Sons, Ltd

Lemma 4.3. Suppose a new example is added to the evidence and it is covered by the theory (*hit*). For each rule r that is used for it, the new $\rho'(r)$ can be easily obtained from the old $\rho(r)$ by:

$$\rho'(r) = [\,\rho(r) + 1\,] / 2$$

Proof. The new $\rho\rho'(r)$ is incremented by one, i.e. $\rho\rho'(r)=\rho\rho(r) + 1$. From here, $\rho'(r)=1 - 2^{-\rho\rho'(r)} = 1 - 2^{-\rho\rho(r)-1} = 1 - 2^{-\rho\rho(r)}/2 = \frac{1}{2} \cdot [2 - 2^{-\rho\rho(r)}] = \frac{1}{2} \cdot [\,1 + 1 - 2^{-\rho\rho(r)}\,] = \frac{1}{2} \cdot [\,1 + \rho(r)]$. □

Corollary. If an example is removed from the evidence, for each rule r that was used for it, the new $\rho'(r)$ can be easily obtained from the old $\rho(r)$ by:

$$\rho'(r) = 2 \cdot \rho(r) - 1$$

Hence, if a rule r covers a single example then $\rho(r)=0.5$ and if r becomes not necessary, $\rho'(r)=0$.

Definition 4.4. The mean reinforced ratio $m\rho(T)$ is defined as

$$m\rho(T) = \Sigma_{r\in T}\, \rho(r)/m,$$

being m the number of rules.

From these definitions one can verify that, in general, the most (*mean*) reinforced theory is not the shortest one as the following example shows:

EXAMPLE 4.1

Given the evidence e_1,e_2,e_3, consider a theory $T_a=\{r_1,r_2,r_3\}$ where $\{r_1\}$ covers $\{e_1\}$, $\{r_2\}$ covers $\{e_2\}$ and $\{r_3\}$ covers $\{e_3\}$ and a theory $T_b=\{r_1,r_2,r_3,r_4\}$ where $\{r_1,r_4\}$ cover $\{e_1\}$, $\{r_2,r_4\}$ cover $\{e_2\}$ and $\{r_3,r_4\}$ cover $\{e_3\}$.

From here, T_a is less reinforced than T_b.

In the first case we have $\rho\rho_{a,1}= \rho\rho_{a,2}= \rho\rho_{a,3}= 1$ and $m\rho(T_a) = 0.5$. For T_b we have $\rho\rho_{b,1}= \rho\rho_{b,2}= \rho\rho_{b,3}= 1$, $\rho\rho_{b,4}= 3$ and $m\rho(T_b) = 0.5938$.

In addition, redundancy does not imply a loss of mean reinforcement ratio (e.g. just add twice the same rule). However, this measurement of the *theory* allows *fantastic* (unreal) concepts:

Theorem 4.5. Consider a program P composed of rules r_i of the form $\{\, h :- t_1, t_2, .. t_s \,\}$, which covers n examples $E = \{\, e_1, e_2, ... e_n \,\}$. If the mean reinforcement ratio $m\rho < 1 - 2^{-n}$ then it can always be increased.

Proof. A *fantastic* rule r_f can be added to the program by modifying all the rules of the program in the following way $r_i = \{\, h :- t_1, t_2, .. t_s , r_f \,\}$. Obviously, all the other rules maintain the same reinforcement but r_f is now reinforced with $\rho\rho(r_f) = n$. Since $\rho(r_f) > m\rho$ then the new $m\rho'$ must be greater than $m\rho$. □

One can argue that these *fantastic* rules could be checked out and eliminated. However, there are many ways to 'hide' a fantastic rule; in fact, cryptography relies on this fact.

5. REINFORCEMENT WRT. THE EVIDENCE

It might be derived from this problem that reinforcement must be combined with a simplicity criterion in order to work. There is an analogy with neural networks, where this avoidance of overfitting, ensured by simplicity, has been more thoroughfully studied in combination with reinforcement. However, the analogy inspires a solution without explicitly making use of simplicity. The idea is to measure the validation *wrt. the evidence*.

Definition 5.1. The course $\chi_T(f)$ of a given fact f wrt. a theory T is defined as:

$$\chi_T(f) = max_{S\subset Proof(e,T)}\, \{\, \Pi_{r\in S}\, \rho(r) \,\}$$

More constructively, $\chi_T(f)$ is computed as the product of all the reinforcements $\rho(r)$ of all the rules r of T used in the proof of f. If a rule is used more than once, it is computed once. If f has more than one proof, we select the greatest course.

© 1999 John Wiley & Sons, Ltd

The way reinforcements are calculated avoids the generation of very complex programs, but redundancy is possible. However, now there is not any risk of fantastic concepts. As said before, for any program P composed of rules r_i of the form { h :- t_1, t_2, .. t_s }, which covers m examples E = { e_1, e_2, ... e_n } and their reinforcements ρ_i, a *fantastic* rule r_f could be added to the program and all the rules could be modified in the following way r_i = { h :- t_1, t_2, .. t_s, r_f }. The following theorem shows that now it is not reinforced over the original one:

Theorem 5.2. The course of any example cannot be increased by the use of *fantastic* concepts.

Proof. Since the *fantastic* concept r_f now appears in all the proofs of the n examples, the reinforcement of r_f is exactly $1 - 2^{-n}$ and the reinforcements of all the r_i remain the same. Hence, the course of all the n examples is modified to $\chi'(e_j) = \chi(e_j) \cdot r_f = \chi(e_j) - \chi(e_j) \cdot 2^{-n}$. Since n is finite, for all $e_j \in E$, $\chi'(e_j)$ can never be greater than $\chi(e_j)$.□

6. APPLICATIONS

Now it is time to start to use the previous measure to evaluate inductive theories. The first idea is to use the greatest *mean* of the courses of all the data presented so far, defined as:

Definition 6.1. The mean course $m\chi(T, E)$ of a theory T wrt. an evidence E is defined as:

$$m\chi(T, E) = \Sigma_{e \in E}\, \chi_T(e) / n$$

being n = card(E).

In order to obtain a more compensated theory, a geometric mean can be used instead, which we will denote by $\mu\chi$. For every theory T, we will say that it is *worthy* for E iff $m\chi(T, E) \geq 0.5$. If the representation language is expressible enough, it is easy to show that for every evidence E there is at least a theory worthy for it (just choose a theory with an extensional rule for covering each example). The same holds for $\mu\chi$.

6.1. Knowledge Construction, Revision and Abduction

The use of these simple measurements can be seen in the following example, which is somehow *long* in order to show the use of this *new* criterion for knowledge construction:

EXAMPLE 6.1

Using Horn theories for representation (Prolog), consider the following incremental learning session:

⌦ Given the background theory B = { s(a,b), s(b,c), s(c,d) } we observe the evidence

E = { e^+_1: r(a,b,c), e^+_2: r(b,c,d), e^+_3: r(a,c,d), e^-_1: ¬r(b,a,c), e^-_2: ¬r(c,a,c) }:

The following programs could be induced, with their corresponding reinforcements and courses:

P_1 = {r(X,Y,Z) :- s(Y,Z) : ρ = 0.875} — $\chi(e^+_1)$= $\chi(e^+_2)$= $\chi(e^+_3)$= 0.875

P_2 = {r(X,c,Z) : ρ = 0.75
r(a,Y,Z) : ρ = 0.75} — $\chi(e^+_1)$= $\chi(e^+_2)$= $\chi(e^+_3)$= 0.75

P_3 = {r(X,Y,Z) :- s(X,Y) : ρ = 0.75
r(X,Y,Z) :- s(Y,Z) : ρ = 0.875} — $\chi(e^+_1)$= $\chi(e^+_2)$= $\chi(e^+_3)$= 0.875

P_4 = {r(X,Y,Z) :- t(X,Y), t(Y,Z) : ρ = 0.875
t(X,Y) :- s(X,Y) : ρ = 0.875
t(X,Y) :- s(X,Z), t(Z,Y) : ρ = 0.5} — $\chi(e^+_1)$= $\chi(e^+_2)$= 0.7656, $\chi(e^+_3)$= 0.3828

P_5 = {r(X,Y,Z) :- t(X,Y) : ρ = 0.875
t(X,Y) :- s(X,Y) : ρ = 0.875
t(X,Y) :- s(X,Z), t(Z,Y) : ρ = 0.5} — $\chi(e^+_1)$= $\chi(e^+_2)$= 0.7656, $\chi(e^+_3)$= 0.3828

At this moment, P_1 and P_3 are the best options and P_4 and P_5 seem 'risky' theories wrt. the evidence.

⌦ e^+_4 = r(a,b,d) is observed.

P_1 does not cover e_4^+ and it is patched:

P_{1a}' = {r(X,Y,Z) :- s(Y,Z) : ρ = 0.875 — $\chi(e^+_1)= \chi(e^+_2)= \chi(e^+_3) = 0.875$, $\chi(e^+_4) = 0.5$
r(a,b,d) : ρ = 0.5} — mχ = 0.78, μχ = 0.76

P_{1b}' = {r(X,Y,Z) :- s(Y,Z) : ρ = 0.875
r(X,Y,d) : ρ = 0.875 } — $\chi(e^+_1)= \chi(e^+_2)= \chi(e^+_3)= \chi(e^+_4) = 0.875$

P_2' is reinforced = {r(X,c,Z) : ρ = 0.75.
r(a,Y,Z) : ρ = 0.875} — $\chi(e^+_1)$=0.875, $\chi(e^+_2)$=0.75, $\chi(e^+_3)$=$\chi(e^+_4)$=0.875

P_3' is reinforced = {r(X,Y,Z) :- s(X,Y) : ρ = 0.875.
r(X,Y,Z) :- s(Y,Z) : ρ = 0.875} — $\chi(e^+_1)= \chi(e^+_2)= \chi(e^+_3)= \chi(e^+_4) = 0.875$

P_4' is reinforced

P_4'= { r(X,Y,Z):-t(X,Y), t(Y,Z): ρ = 0.9375
t(X,Y) :- s(X,Y) : ρ = 0.9375 — $\chi(e^+_1)= \chi(e^+_2)= 0.8789$, $\chi(e^+_3)= \chi(e^+_4) = 0.6592$
t(X,Y) :- s(X,Z), t(Z,Y) : ρ = 0.75} — mχ = 0.77, μχ= 0.76

P_5' is slightly reinforced

P_5' = { r(X,Y,Z) :- t(X,Y) : ρ = 0.9375. — $\chi(e^+_1)=\chi(e^+_2)=0.8789$
t(X,Y) :- s(X,Y) : ρ = 0.9375 — $\chi(e^+_3)=0.4395$, $\chi(e^+_4)=0.8789$
t(X,Y) :- s(X,Z), t(Z,Y) : ρ = 0.5} — mχ = 0.77, μχ= 0.74

At this moment, P_{1b}' and P_3' are the best options. Now P_4' and P_5' seem more grounded.

⌦ We add e^-_3 = ¬r(a,d,d)

P_{1a}' remains the same and P_{1b}' and P_2' are inconsistent, motivating the following 'patches' for them:

P_{2a}' = {r(X,c,Z) : ρ = 0.75.
r(X,b,Z) : ρ = 0.75} — $\chi(e^+_1)= \chi(e^+_2)= \chi(e^+_3)= \chi(e^+_4) = 0.75$

P_{2b}' = {r(X,Y,Z) :- e(Y) : ρ = 0.9375.
e(b) : ρ = 0.75
e(c) : ρ = 0.75} — $\chi(e^+_1)= \chi(e^+_2)= \chi(e^+_3)= \chi(e^+_4) = 0.7031$

P_3' and P_4' remain the same. P_5' becomes inconsistent.

⌦ We add e^+_5 = r(a,d,e)

P_{1a}', P_{2a}', P_{2b}' can only be patched with e^+_5 as an exception because abduction is not possible.

P_3' has abduction as a better option.

P_3" = {s(d,e) : ρ = 0.5 — $\chi(e^+_1)=\chi(e^+_2)=\chi(e^+_3)= 0.9375$
r(X,Y,Z) :- s(X,Y) : ρ = 0.875 — $\chi(e^+_4)= 0.875$, $\chi(e^+_5)= 0.4688$
r(X,Y,Z) :- s(Y,Z) : ρ = 0.9375} — mχ = 0.831, μχ= 0.805

P_4' makes the same abduction

P_4" = { s(d,e) : ρ = 0.5
r(X,Y,Z):-t(X,Y),t(Y,Z): ρ=0.96875 — $\chi(e^+_1)=\chi(e^+_2)=0.939$
t(X,Y) :- s(X,Y) : ρ = 0.96875 — $\chi(e^+_3)=\chi(e^+_4)=0.82$,$\chi(e^+_5)=0.41$
t(X,Y) :- s(X,Z), t(Z,Y): ρ = 0.875} — mχ = 0.786, μχ= 0.754

At this moment, P_3" and P_4" are the best options.

Further examples would be required to distinguish with more reliability which is the 'intended' one.

The example illustrates that, in general, and by using this new recknoning of reinforcement, the shortest theories are not the best ones. More importantly, the weak parts are detected by a low value of reinforcement, and revision, if necessary, should be done to these parts of the theory. On the other hand, as soon as a theory gains some solidity, in terms of increase of reinforcement, abduction can be applied. Another advantage of this approach is that a 'rated' ontology can be derived directly from the theory.

© 1999 John Wiley & Sons, Ltd

6.2. Consilience can be precisely defined

We have previously commented on the difficulty of determining in a formal way the idea of 'consilience', introduced by Whewell in the last century[48], and other related concepts, like Reichenbach's principle of common cause, Thagard's coherence[44], all of them with the common idea of giving a conciliating theory for all the data, i.e., all the evidence must be accounted by the same explanation or by very close related explanations.

In the context of reinforcement, it is easy to define consilience:

Definition 6.2. A theory T is partitionable wrt. an evidence E iff $\exists T_1, T_2 : T_1 \subset T$, $T_2 \subset T$ and $T_1 \neq T_2$ such that $\forall e \in E : T_1 \models e \vee T_2 \models e$. We define $E_1 = \{ e \in E : T_1 \models e \}$ and $E_2 = \{ e \in E : T_2 \models e \}$ and $E_{12} = E_1 \cap E_2$. Finally, we will use ther term $S\chi(T_1 \oplus T_2, E)$ to denote the expression $m\chi(T_1, E_1) \cdot [\,\mathrm{card}(E_1) - \mathrm{card}(E_{12})/2 + m\chi(T_2, E_2) \cdot [\,\mathrm{card}(E_2) - \mathrm{card}(E_{12})/2$.

Definition 6.3. A theory T is consilient wrt. an evidence E iff there does not exist a partition T_1, T_2 such that $S\chi(T_1 \oplus T_2, E) \geq m\chi(T, E) \cdot \mathrm{card}(E)$.

In other words, a theory T is consilient wrt. an evidence E iff there does not exist a bipartition $P \in \wp(T)$, such that every example of E is still covered separatedly without loss of reinforcement.

EXAMPLE 6.1

Given the following evidence (in Prolog):

E = { p(a), p(b), p(e), q(a), q(b), q(e), q(f) }

The following program could be induced, with its corresponding reinforcements and courses:

P = { p(X) : ρ = 0.875
q(X) : ρ = 0.9375} $\quad m\chi(E, P) = 0.9107$

The following partition:

P_1 = { p(X) : ρ = 0.875 } $\quad m\chi(E_1, P_1) = 0.875$

P_2 = { q(X) : ρ = 0.9375} $\quad m\chi(E_2, P_2) = 0.9375$

In this case it is obvious that $S\chi(P_1 \oplus P_2, E) = m\chi(E_1, P_1) \cdot 3 + m\chi(E_2, P_2) \cdot 4 = m\chi(E, P) \cdot \mathrm{card}(E) = 0.9107 \cdot 7$. Hence, as expected, P is not consilient.

The definition can be parametrised introducing a consilience factor or changing the arithmetic mean by the geometric mean.

6.3. Intrinsical Exceptions, Consilience and Noise

Using reinforcement, an intrinsecal exception or extensional patch can be easily defined as a rule r with $\rho = 0.5$, i.e. a rule that just covers one example e, or, in other words, it is necessary for only one example. However we must distinguish between completely extensional exceptions, when r does not use any rule from the theory to cover e, and partially extensional exceptions when r uses other rules to describe e. The following theorem shows that completely extensional exceptions should be avoided to obtain consilient programs.

Theorem 6.4. If a worthy theory T for an evidence E has a rule r with $\rho = 0.5$, and completely extensional, then T is not consilient.

Proof. Just choose the partition $T_1 = T - r$ and $T_2 = T$. Since $\rho = 0.5$ then r is only used by one example e_r. Since it is a completely extensional exception, we have that r does not use any rule from T_1 to cover e_r, so $\rho'(r_i) = \rho(r_i)$ for all $r_i \in T_1$. Let n be the number of the examples of the evidence E. Hence, $m\chi(T_1, E_1) = [m\chi(T, E) \cdot n - \chi(e_r, T)\,] \,/\, (n-1) = [m\chi(T, E) \cdot n - ½\,] \,/\, (n-1) = [m\chi(T, E) \cdot n + m\chi(T, E) - m\chi(T, E) - ½\,] \,/\, (n-1) = m\chi(T, E) + [m\chi(T, E) - ½] \,/\, (n-1)$.

From definition 6.3, the disequality simplifies as follows:

$$S\chi(T_1 \oplus T_2, E) =$$

$$m\chi(T_1, E_1) \cdot [\,\mathrm{card}(E_1) - \mathrm{card}(E_{12})/2\,] + m\chi(T_2, E_2) \cdot [\,\mathrm{card}(E_2) - \mathrm{card}(E_{12})/2\,] =$$

$$\{ m\chi(T, E) + [m\chi(T, E) - ½] / (n-1) \} \cdot [(n-1) - (n-1)/2] + m\chi(T, E) \cdot [n - (n-1)/2] =$$
$$m\chi(T, E) \cdot [(n-1) - (n-1)/2 + n - (n-1)/2] + [m\chi(T, E) - ½] \cdot [(n-1) - (n-1)/2] / (n-1) =$$
$$m\chi(T, E) \cdot [n] + [m\chi(T, E) - ½] / 2$$

Since T is worthy, then $m\chi(T, E) \geq 0.5$., and finally

$$S\chi(T_1 \oplus T_2, E) \geq m\chi(T, E) \cdot n = m\chi(T, E) \cdot \text{card}(E). \square$$

In the same way, partially extensional exceptions are not suitable for consilience, but a limit would depend on how many rules are been used by the exception, because the separation would make the reinforcement of these rules decrease as follows $\rho'(r_i)=2\cdot\rho(r_i)-1$, by the corollary of lemma 4.3.

In any case, not only intensionality (avoidance of exceptions) but consilience are both very strict requirements in the presence of noise, because any piece of data which is left as noise would be tried to be 'conciled' with the rest of the theory, sometimes in an artificial way.

However, if used correctly, reinforcement is a very powerful tool to control the level of noise in a theory. This means that if we have any information or hint about the expected noise ratio, we can adjust the percentage of examples covered by extensional rules.

7. REINFORCEMENT, INTENSIONALITY AND CROSS-VALIDATION

The idea of intensionality is useful to distinguish between explanatory views of induction (and abduction as a particular case) and non-explanatory induction[16], where the goal is to describe compactly the evidence, but not to explain *all* of it. Moreover, there is a strong relation between intensionality (or avoidance of exceptions) and hypothesis stability.

In this section we will make the connection between intensionality (i.e. avoidance of exceptions, as they were defined in the previous section) and cross-validation. Since Devroye and Wagner[7] established the relation between leave-one-out cross-validation and hypothesis stability, many other variants of cross-validation have been studied (like training-test split or *k*-fold).

We will work with multi-fold split, that is to say, we will take into account all the possible splits in all the possible orders, to see the influence of intrinsical exceptions in the stability of the theory. Let us denote with n_e the number of rules r that just cover one example e. In other words, if the example e would have not appeared, the rule r would be useless. We will make the following reasonable assumption: a natural learning algorithm is a learning algorithm that does not add useless rules to the theory.

Let us define $P(\mathcal{A},T,E,k)$ as the probability that the algorithm $\mathcal{A}$ gives the theory T with the first k examples of the evidence E, considering all possible orderings of E.

Theorem 7.1. For any natural learning algorithm $\mathcal{A}$,

$$P(\mathcal{A},T,E,k) \leq 1 - [(n-n_e)^{n-k} / n^{n-k}]$$

being $n = \text{card}(E)$.

Proof. Let us denote with E^w the examples from E that are covered by a rule with $\rho = 0.5$. Let $w = \text{card}(E^w)$, $E^b = E - E^w$ and $b = n-w$. Obviously, $w \leq n_e$ since there can be examples covered by more than one exception rule. We denote with $E^{1..k}$ and $E^{k+1..n}$ the set of the first k examples and the rest of the n examples of a given ordering of E, in other words, a split at position k. We define $P^w(E,k)$ as the probability of $E^w \cap E^{k+1..n} \neq \emptyset$, i.e., the probability of having one exception example in the second part of the split. By a simple combinatorial analysis, removing from the whole probability the probability of having all $E^{k+1..n}$ from E^b, this probability is:

$$P^w(E,k) = [(w + b)^{n-k} - b^{n-k}] / n^{n-k} = 1 - [b^{n-k} / n^{n-k}]$$

Since $b = n-w$, we have

$$P^w(E,k) = 1 - [\,(n-w)^{n-k} \;/\; n^{n-k}\,]$$

and $w \leq n_e$, so

$$P^w(E,k) \leq 1 - [\,(n-n_e)^{n-k} \;/\; n^{n-k}\,]$$

From here, $P(\mathcal{A},T,E,k) \leq P^w(E,k)$ because $\mathcal{A}$ is natural.□

At first sight the result may be understood as a rationale to avoid exceptions, in order to have $P^w(E,k)=1$. For instance, given a theory of 100 rules, 3 of which are exception rules, we have that the probability that the theory could be found with eighty examples is $P^w(E, 80) \leq 1-97^{20}/100^{20}=0.45$.

The ideas of intensionality have been used in an incremental learning system[18] using Curry as a representation language (a logic functional programming language based on narrowing with some higher-order constructs). The results demonstrate that the *intended* hypothesis is found sooner than when the MDL principle is used, because the latter allows the introduction of patches (exceptions) in an incremental session.

A deeper reflection on theorem 7.1 shows that stability *of the whole theory* is a very strict requirement. If it is substituted by partial stability, i.e., how many rules of the theory can be obtained in early learning steps, the result may be quite different. Moreover, the connection between mean course and cross-validation would be more enlightening, although more difficult to obtain.

In the end, theorem 7.1 is just an example of the connections that could be established between model selection methods for constructive languages, by using reinforcement as a measure. In this section it has been done with a particular variant of cross-validation. However, these connections can be established at a higher and more general level than other comparisons based on error estimation and attribute complexities[22]. The next section will address the relation with the MDL principle.

8. BALANCED REINFORCEMENT

With the final measure introduced in section 5 there is still a tricky way of increasing reinforcement: joining rules. If a high-level representation language allows very expressive rules, larger rules can be made in order to stand for the same that was expressed with separated rules, with the advantage of increasing reinforcement and mean course:

EXAMPLE 8.1

For instance, the following extended functional programs are equivalent:

T_a = { r_1 = { f(X,a) → g(b) },
r_2 = { f(X,c) → i(d) } }

T_b = { r = { f(X,Y) → if (Y=a) then g(b)
if (Y=c) else i(d) } }

but T_b would be more reinforced than T_a.

In order to maintain the granularity of the theory there are two options: (1) the introduction of a factor directly related with the number of rules, and (2) the introduction of a factor inversely related with the syntactical length of each rule. We will choose this second option to clarify that this modification still makes our measure very different from a prior distribution like the MDL principle.

With $\mathrm{length}(r)$ we will denote the length of a rule r for any specific language. The only restriction for length is that for all r, $\mathrm{length}(r) \geq 1$. Thus we extend the definitions of section 5:

Definition 8.1. The extended pure reinforcement is defined as:

$$\rho\rho^*(r) = \rho\rho(r) \;/\; \mathrm{length}(r).$$

The extended normalised reinforcement $\rho^*(r)$ and the extended courses $\chi^*(r)$ are defined in the obvious way using $\rho\rho^*(r)$ and $\rho^*(r)$.

© 1999 John Wiley & Sons, Ltd

With this extension, it is easy to show that —in the limit[12]— the MDL principle is an excellent principle for achieving reinforcement:

Theorem 8.2. If the data E are infinite and a theory T is finite, the mean course $m\chi^*(T, E) = 1$.

Proof. Given some infinite data as evidence $E = \{ e_1, ..., e_n \}$, without loss of generality, consider that T can be exclusively composed of two rules: r_1, which covers all E except e_i and, *independently*, r_2, which covers e_i. The reinforcements are $\rho^*(r_1) = (1-2^{(1-n)/\text{length}(r_1)})$ and $\rho^*(r_2) = (1-2^{-1/\text{length}(r_2)})$ and the mean course $m\chi^*(T, E) = [(n-1) \cdot (1-2^{(1-n)/\text{length}(r_1)}) + (1-2^{-1/\text{length}(r_2)})] / n$. For infinite data, we have that $\lim_{n\to\infty} m\chi^*(T, E) = 1$. □

The result is independent of the last extension given by definition 8.1. In general, the theorem shows that maximum reinforcement matches with maximum compression in the limit (simply because both are saturated). However, when the data are finite we have many cases where they differ. The most blatant case occurs when some exception is covered extensionally (as r_2 which covers d_i in the proof of theorem 8.2) and there is an important loss of reinforcement vs. a slight loss of compression. The following example illustrates this point:

EXAMPLE 8.2

Consider the following evidence e_1–e_{10}:

E = { e_1: e(4) → true, e_2: e(12) → true,
e_3: e(3) → false, e_4: e(2) → true,
e_5: e(7) → false, e_6: e(7) → false,
e_7: e(20) → true, e_8: e(0) → true,
e_9: o(3) → true, e_{10}: o(2) → false }

where natural numbers are represented by using the functor s as the symbol for successor, e.g. s(s(s(0))) means 3. The length (denoted l) of a rule is computed as $1+n_f+n_v$, where n_f means the number of functors (including constants as functors with arity 0) and n_v being the number of variables.

From here, the following theories are evaluated:

	: l	ρρ	ρρ*	ρ*
T_a= { e(s(s(X)) → e(X)	: 7	7	1	0.5
e(0) → true	: 4	5	1.2	0.5647
e(s(0)) → false	: 5	3	0.6	0.3402
o(s(s(s(0)))) → true	: 7	1	0.1429	0.0943
o(s(s(0))) → false	: 6	1	0.1667	0.1091}

The extended courses are $\chi^*(e_1, e_2, e_4, e_7, e_8) = 0.5 \cdot 0.5647 = 0.28235$, $\chi^*(e_3, e_5, e_6) = 0.5 \cdot 0.3402 = 0.1701$, $\chi^*(e_9) = 0.0943$ and $\chi^*(e_{10}) = 0.1091$.

The mean extended course $m\chi^*$' is 0.2125.

	: l	ρρ	ρρ*	ρ*
T_b= { e(s(s(X)) → e(X)	: 7	7	1	0.5
e(0) → true	: 4	5	1.2	0.5647
e(s(0)) → false	: 5	3	0.6	0.3402
o(s(s(X)) → o(X)	: 7	2	0.2857	0.1797
o(0) → false	: 4	1	0.25	0.1591
o(s(0)) → true	: 5	1	0.2	0.1294}

The extended courses are $\chi^*(e_1, e_2, e_4, e_7, e_8) = 0.5 \cdot 0.5647 = 0.28235$, $\chi^*(e_3, e_5, e_6) = 0.5 \cdot 0.3402 = 0.1701$, $\chi^*(e_9) = 0.1797 \cdot 0.1294 = 0.02325$ and $\chi^*(e_{10}) = 0.1797 \cdot 0.1591 = 0.02859$.

The mean extended course $m\chi^*$' is 0.1974.

© 1999 John Wiley & Sons, Ltd

		: l	ρρ	ρρ*	ρ*
T_c= {	e(s(s(X)) → e(X)	: 7	9	1.2857	0.5898
	e(0) → true	: 4	6	1.5	0.6464
	e(s(0)) → false	: 5	4	0.8	0.4257
	o(X) → not(e(X))	: 6	2	0.3333	0.2063
	not(true) → false	: 4	1	0.25	0.1591
	not(false) → true	: 4	1	0.25	0.1591}

The extended courses are $\chi^*(e_1, e_2, e_4, e_7, e_8) = 0.5898 \cdot 0.6464 = 0.3813$, $\chi^*(e_3, e_5, e_6) = 0.5898 \cdot 0.4257 = 0.2511$, $\chi^*(e_9) = 0.2063 \cdot 0.5898 \cdot 0.4257 \cdot 0.1591 = 0.00824$ and $\chi^*(e_{10}) = 0.2063 \cdot 0.5898 \cdot 0.6464 \cdot 0.1591 = 0.0125$.

The mean extended course $m\chi^*$' is 0.2681.

Note that the lengths ($l(T_a)$=29, $l(T_b)$=32, $l(T_c)$ = 30) do not give many hints about which theory to choose, while reinforcement selects more clearly the last one.

The example also shows the advantages of this approach for explanation-based learning. Since all the data should be explained, if a part is left in an extensional way (or unrelated with the rest), it is penalised. On the other hand, we have seen in the preceding sections that *fantastic* concepts are also avoided, so it results to be a *balanced* criterion for a more reasonable degree of theory intensionality, without falling into fantasy.

Regarding T_c of example 8.2, our measure can be adapted to situations where a more compensated theory is required, using a *geometric mean* instead of an *arithmetic mean*. In addition, and concerning T_a, if we do not want exceptions (extensional parts) at all, we discard theories where a fact has a course value less than the mean divided by a constant. In an incremental framework, this case should trigger theory revision in order to integrate (or reconcile) the example with the theory.

9. REWARDED REINFORCEMENT

In reinforcement learning, it is usually assumed that the learner receives some reward (or penalty) value for its actions. In other words, prediction hits receive different degrees of reward. Prediction errors (including novelties and anomalies) receive different degrees of penalty (or negative reward).

Usually, this broader view of reinforcement is suitable for frameworks where reasoning about action is necessary. The rewards are assigned depending on the actions that the agent performs for each situation. Apart from Markov decision processes[23], other more expresible temporal languages are used for representation, like event calculus or situation calculus[25]. The important issue here is that our model selection measures can be used for these high-level representations. The value of reinforcement can be understood as the prediction reliability of the following situation s_{n+1} after every possible action that can be performed in a certain situation s_n. The task of the system seems to be the choice of the one with the greatest reward. With this first approach, in the case the result of the action matches with the evidence, a positive hit happens with the predicted reward. However, in the case a prediction error occurs, the action may have no awful consequences (no penalty) but in some cases, it may be fatal. The question is how ontology and 'hedonism' must be combined. It is commonly accepted in psychology the claim that hedonism motivates ontology, and this is stronger the earlier the stage of development of a cognitive system. In our opinion, this motivation does not imply that they must be mixed. Moreover, rewards should also be learned because they may change.

Hence, the choice of the best action must take into account both the reliability of the prediction (i.e. the reinforcement) weighted with the reward, not the action with the best reward alone (because it may be a very weak guess).

© 1999 John Wiley & Sons, Ltd

Finally, there can be degrees of reliability in the evidence. This degree may come from different reliabilities of the system sensors or from intermediate recognition or sensor preprocessing subsystems. Indeed, this should affect ontology in the following way: every fact of the evidence is assigned a real number as a reliability degree, $-1 \leq d_f \leq 1$. In this framework, the completely reliable positive examples are assigned a value of $d_f = 1$ and the completely reliable negative examples are assigned a value of $d_f = -1$.

Definition 9.1. The *'grounded' course* $\chi'(f)$ of a given fact f wrt. to a theory is computed as the normal course $\chi(f)$ multiplied by the reliability degree of f. More formally, $\chi'(f) = \chi(f) \cdot d_f$.

Summing up, the decision of which action should be taken would depend on:

- the reliability of recognising the situation where the agent is really embedded.
- the reliability of predicting the consequence of a given action in that situation.
- the reward (or penalty) of the consequence (and its reliability if rewards are learned as well).

In the previous section we considered the length of rules. Another straightforward extension to our approach is to consider the length of the examples, too. This can also be incorporated in the same way as the reliability degree.

10. COMPUTING REINFORCEMENT

As the reader must have realised, our theory of reinforcement is not an inductive learning method. We have not dealt about how the theory could be constructed from the evidence. On the contrary, this paper have presented a setting for constructive reinforcement learning based on a measurement that allows a detailed study of the relation between the theory and the evidence, for assisting the evaluation, the selection, and the revision of theories.

Notwithstanding, the measurement needs to be computed. A general method of computing reinforcement is just as it has been used in all the examples which have appeared throughout the paper:

GENERAL METHOD:

Consider the theory T, with m rules $r_1..r_m$, and the evidence E, with n examples $e_1..e_n$, such that $T \models E$. First we must *prove* all the examples and compute $\rho\rho^*$ and ρ^* for each rule. In a second stage, we *prove* again the n examples, computing χ^* from the ρ^* obtained in the first stage.

The complexity of the previous method *seems* to be, in the worst case, in $O(m \cdot n)$. However it is not so, because we have not stated any restriction about the computational cost of the theory, and each proof has its own cost. Nonetheless, it would be more realistic to consider the reckoning of reinforcement in an incremental setting:

INCREMENTAL METHOD:

We will use four arrays: $l_{1..m}$, $\rho\rho^*_{1..m}$, $\rho^*_{1..m}$, $\chi^*_{1..n}$ for the lengths, the pure and normalised reinforcements and the courses, respectively. An additional boolean bidimensional array $U_{1..m,1..n}$ assigns *true* to $U_{j,i}$ iff e_i uses r_m in its proof and *false* otherwise.

For each new example e_{n+1} which is received we have different possibilities:

1. If it is a *hit*, we remake $\rho\rho^*_{1..m}$, $\rho^*_{1..m}$, according to the proof of e_{n+1}, U is extended to $U_{.n+1}$ and $\chi^*_{1..n+1}$ is updated using U.
2. If it is a *novelty* and no revision is made to T, only an extension $T' = T \cup \{r_{m+1}, ..., r_{m+k}\}$, the steps are very similiar to the previous case, except that the arrays must be extended to $m+k$.
3. Finally, if it is a *novelty* or an *anomaly* and the theory is revised in some rules $\{r_1, ..., r_p\}$ and extended in others $\{r_{m+1}, ..., r_{m+k}\}$, only the $U_{.j}$ which does not use any rule from $\{r_1, ..., r_p\}$ can be preserved. The rest must be remade.

© 1999 John Wiley & Sons, Ltd

The previous method ignores two exceptional cases: that a *hit* would trigger a revision of the theory to readjust reinforcements, and that case 2. may produce alternative proofs for previous examples.

Further optimisation could come from a deeper study of the static dependencies (i.e. some rule always depends on others) and the topology of the dependencies that the theory generates. On the other hand, an appropriate approximation could be used. Even more, part of the past evidence can be 'forgotten' if it is covered by very reinforced rules, so avoiding future computations to a large extent without a significant loss in accuracy.

However, in the case that an inductive learning method uses reinforcement for evaluating the theories it is constructing, the complexity of these methods would surely be very modest compared to the usual huge costs of machine learning algorithms.

Moreover, reinforcement measures are a very adequate tool to guide a learning method. For instance, in a learning algorithm for logic functional languages based on genetic programming[18], the examples and rules with low reinforcement are mixed first in order to 'conciliate' them into more compact and reinforced theories.

11. CONCLUSIONS

This paper introduces a mechanism for propagating reinforcement into constructive theories depending on the observation (or evidence). Strictly speaking, this model selection criterion is neither syntactical nor semantical. It is, as we have dubbed, structural, i.e., it is based on "how the hypothesis cover the evidence". The advantage of this approach is that it makes no assumptions about the prior distribution. Also in this framework, knowledge can have alternative descriptions, without reducing the evidence's courses. Moreover, "deduction in the knowledge" is possible and it can even affect positively to reinforcement. In contrast, the MDL principle or other syntactical prior selection methods cannot use deductive inference without decreasing the a posteriori probability.

Reinforcement allows a more detailed treatment of exceptions and provides different ratings for different parts of a theory, not the single probability value given by the prior distribution which is assigned to the whole theory. Moreover, different predictions or assumptions are provided with different reliability values.

We have presented some examples, using logical and functional languages. They illustrate the utility of our framework in the context of knowledge construction. They also show that abduction is feasible as long as the theory gets reinforced. Other reasoning processes like analogy can also be ellucidated under this view of reinforcement. Although directly applicable to expert systems, diagnostic systems and ILP frameworks, conceptually, this work is closer to the distribution of reinforcement in neural networks (training = induction, recognition = abduction), and the problems of overfitting and underfitting. It even resembles some popular algorithms, like back-propagation. However, a symbolical framework with topological flexibility allows the direct combination with different areas and applications which have used or may use it in the future: ILP, EBL, Analogical Reasoning, Reinforcement Learning and some kinds of non-monotonic reasoning, much more knowledge oriented than artificial neural networks or other hybrid approaches.

At present, it is more compelling to continue the evaluation of our measures in practice[18]. As future work, the measurement could be extended to consider time-complexity and/or negative cases in the courses. In addition, a deeper study of how deduction affects reinforcement could be of capital interest in knowledge-based systems which use inductive and deductive reasoning techniques. Finally, we plan to apply our ideas in domains with actions, probably using situation or event calculus[25,40], and to treat rewards in a more direct way than it has been done in section 9 (connecting with

© 1999 John Wiley & Sons, Ltd

the work of Dietterich & Flann[8]), in order to re-associate our notion of reinforcement with more classical notions of reinforcement learning.

ACKNOWLEDGEMENTS

The author would like to thank the referees of CCIA'98 for their comments on the initial version of this work, especially for suggesting the improvement of section 9 and the introduction of section 10. The proof of theorem 7.1 is partially due to Neus Minaya. The author is also grateful to Boris D. Siepert for several corrections and general comments about earlier drafts of this paper.

REFERENCES

1. D.W. Aha, "Lazy Learning. Editorial" Special Issue about "Lazy Learning" AI Review, v.11, **1-5**, Feb. (1997).
2. A. Aliseda "A Unified Framework for Abductive and Inductive Reasoning in Philosophy and AI" in M. Denecker, L. De Raedt, P. Flach and T. Kakas (eds.) ECAI'96 Workshop on *Abductive and Inductive Reasoning*, pp. 7-9, 1996.
3. S.F. Barker *Induction and Hypothesis* Ithaca, 1957.
4. A. Blumer, A. Ehrenfeucht, D. Haussler and M.K. Warmuth "Occam's razor" *Inf. Proc. Letters*, **24**, 377-380 (1987).
5. C. Botilier and V. Becher, "Abduction as belief revision" *Artificial Intelligence* **77**, 43-94, (1995).
6. T. Bylander, M.C. Allemang, M.C. Tanner and J.R. Josephson, "The computational complexity of abduction" *Artificial Intelligence*, **49**, 25-60, (1991).
7. L.P. Devroye and T.J. Wagner, "Distribution-free performance bounds for potential function rules" *IEEE Transactions on Information Theory*, IT-**25**(5):601-604, 1979.
8. T.G. Dietterich and N.S. Flann, "Explanation-Based Learning and Reinforcement Learning: A Unified View" *Machine Learning*, **28**, 169-210, (1997).
9. R. Ernis, "Enumerative Induction and Best Explanation" *J. of Philosophy*, **LXV** (18), 523-529, (1968).
10. P. Flach, "Abduction and Induction: Syllogistic and Inferential Perspectives" in M. Denecker, L. De Raedt, P. Flach and T. Kakas (eds) *Working Notes of the ECAI'96 Workshop on Abductive and Inductive Reasoning*, pp. 7-9, 1996.
11. P. Flach and A. Kakas (eds.), *Abduction and Induction. Essays on their relation and integration*, in press, Kluwer.
12. E.M. Gold, "Language Identification in the Limit" *Inform. and Control*., **10**, pp. 447-474, (1967).
13. P. Grünwald, "The Minimum Description Length Principle and Non-Deductive Inference" in P. Flach and A. Kakas (eds.), Proceedings of the IJCAI'97 Workshop on *Abduction and Induction in AI*, Nagoya, Japan 1997.
14. G. Harman, "The inference to the best explanation" *Philosophical Review*, **74**, 88-95, (1965).
15. C.G. Hempel, *Aspects of Scientific Explanation*, The Free Press, New York, N.Y. 1965.
16. J. Hernández-Orallo and I. García-Varea, "Distinguishing Abduction and Induction under Intensional Complexity" in P. Flach and A. Kakas (eds.) *Proc. of the European Conference of Artificial Intelligence (ECAI'98) Ws. on Abduction and Induction in AI*, pp. 41-48, Brighton 1998.
17. J. Hernández-Orallo and I. García-Varea, "On Autistic Interpretations of Occam's Razor", http://www.dsic.upv.es/~jorallo/escrits/autistic21.ps.gz, to be presented at the Intl. Conf. on Model Based Reasoning (MBR'98), Pavia, Italy.
18. J. Hernández-Orallo and M.J. Ramírez-Quintana, "Inductive Inference of Functional Logic Programs by Inverse Narrowing" J. Lloyd (ed) *Proc. JICSLP'98 CompulogNet Meeting on Comp. Logic & Machine Learning*, pp. 49-55, 1998.
19. J.H. Holland, K.J. Holyoak, R.E. Nisbett and P.R. Thagard, *Induction, Processes of Inference, Learning and Discovery*, The MIT Press, 1986.
20. L. Kaelbling, M. Littman and A. Moore, "Reinforcement Learning: A survey" *J. of AI Research*, **4**, 237-285, (1996).
21. A. Karmiloff-Smith, *Beyond Modularity: A Developmental Prespective on Cognitive Science*, The MIT Press 1992.
22. M. Kearns, Y. Mansour, A.Y. Ng and D. Ron, "An Experimental and Theoretical Comparison of Model Selection Methods" *Machine Learning*, to appear. URL: http://www.research.att.com/~mkearns/
23. M. Kearns and S. Singh, "Near-Optimal Performance for Reinforcement Learning in Polynomial Time" URL: http://www.research.att.com/~mkearns/
24. A.N. Kolmogorov, "Three Approaches to the Quantitative Definition of Information" *Problems Inform. Transmission*, **1**(1): 1-7, (1965).
25. R. Kowalski and F. Sachi "Reconciling the Event Calculus with the Situation Calculus" *J.Logic Prog.* **31**(1-3), 39-58, (1997).
26. T.S. Kuhn, *The Structure of Scientific Revolution*, University of Chigago 1970.
27. L.A. Levin "Universal search problems" *Problems Inform. Transmission*, **9**, 265-266, (1973).
28. R. Levinson "General game-playing and reinforcement learning" *Computational Intelligence*, **12**(1): 155-176, (1996).

© 1999 John Wiley & Sons, Ltd

29. M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and its Applications*, 2nd Ed. Springer-Verlag 1997.
30. R. López de Mántaras and E. Armengol, "Machine Learning from examples: Inductive and Lazy Methods" *Data & Knowledge Engineering* **25**, 99-123, (1998).
31. R.S. Michalski, "Concept Learning" in S.C. Shapiro (ed). *Encyclopedia of AI*, 185-194, John Wiley , 1987.
32. T.M. Mitchell, *Machine Learning*, McGraw-Hill Series in Computer Science, 1997.
33. R.J. Mooney "Integrating Abduction and Induction in Machine Learning" in Peter Flach and Antonis Kakas (eds), *Proceedings of the IJCAI'97 Workshop on Abduction and Induction in AI*, Nagoya, Japan 1997.
34. S. Muggleton and L. De Raedt, "Inductive Logic Programming — theory and methods" *J.Logic Prog.*, **19-20**, 629-679, (1994).
35. S. Muggleton, and C.D. Page, "A Learnability Model for Universal Representations" Unpublished Manuscript, Oxford University Computing Laboratory, 1995. URL: http://www.cs.york.ac.uk/~stephen/jnl.html
36. K.R. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*, Basic Books, 1962.
37. J. Rissanen, "Modelling by the shortest data description" *Automatica-J.IFAC*, **14**, 465-471, (1978).
38. J. Rissanen, "Fisher Information and Stochastic Complexity" *IEEE Trans. Inf. Theory*, 1(**42**): 40-47, (1996).
39. J. Schmidhuber, J. Zhao and M. Wiering, "Shifting Inductive Bias with Success-Story Algorithm, Adaptive Levin Search, and Incremental Self-Improvement" *Machine Learning*, **28**, 105-132, (1997).
40. M. Shanahan, "Explanation in the Situation Calculus" Proceedings of IJCAI'93, pp. 160-165, 1993.
41. E. Shapiro, "Inductive Inference of Theories form Facts" Research Report 192, Dep. of Computer Science, Yale Univ., 1981, also in Lassez, J.; Plotking, G. (eds.) *Computational Logic,* The MIT Press 1991.
42. R.J. Solomonoff, "A formal theory of inductive inference" *Inf. Control* v.**7**, 1-22, Mar., 224-254, June (1964).
43. R.S. Sutton, "Special issue on reinforcement learning" *Machine Learning*, 1991.
44. P.Thagard, "The best explanation: Criteria for theory choice" *Journal of Philosophy*, **75**, 76-.92 (1978).
45. L. Valiant, "A theory of the learnable" *Communication of the ACM*, **27** (11), pp. 1134-1142, 1984.
46. van den Bosch, *Simplicity and Prediction*, Master Thesis, Dep. of Science, Logic & Epistemology of the Fac. of Philosophy at the Univ. of Groningen, 1994.
47. P.Vitányi and M. Li, "On Prediction by Data Compression", *Proc. 9th European Conference on Machine Learning*, Lecture Notes in AI, Vol. 1224, Springer-Verlag, 14-30, 1997.
48. W. Whewell, "The Philosophy of the Inductive Sciences" New York: Johnson Reprint Corp, 1847.

© 1999 John Wiley & Sons, Ltd