INPUT "UNDERSTANDING" AS A BASIS FOR MULTISTRATEGY TASK-ADAPTIVE LEARNING

by

R. S. Michalski
G. Tecuci

Proceedings of the International Symposium on Methodologies for Intelligent Systems
Lecture Notes on Artificial Intelligence, Springer Verlag, 1991.

in Z.Ras and M.Zemankova (eds), Methodologies for Intelligent Systems,
Springer Verlag, 1991

INPUT UNDERSTANDING AS A BASIS FOR MULTISTRATEGY TASK-ADAPTIVE LEARNING

Gheorghe Tecuci* and Ryszard S. Michalski
Center for Artificial Intelligence and Department of Computer Science
George Mason University, 4400 University Drive
Fairfax, VA 22030-4444, USA

Abstract

The paper explores several general issues in developing a multistrategy task-adaptive learning (MTL) system. The system aims at integrating a whole range of learning strategies, such as explanation-based learning, empirical generalization, abduction, constructive induction, learning by analogy and abstraction. The integration is dynamic, i.e. the way different strategies are evoked depends on the learning task at hand. The key idea of the learning method is that the learner tries to "understand" the input in terms of its current knowledge, and then uses this understanding to improve the knowledge. This process may involve both certain and plausible reasoning. The paper extends and generalizes the previous work on this topic.

Keywords: multistrategy learning, induction, analogy, abduction, abstraction, explanation-based learning, knowledge acquisition

1 Introduction

The rapidly evolving field of machine learning increasingly emphasizes the need for addressing real-world learning problems. The nature and complexity of such problems often makes invalid the assumptions needed for applying a single strategy method, such as empirical induction, explanation-based learning, learning by analogy, case-based reasoning, or abductive learning. A striking feature, however, of these single-strategy learning methods is the complementary nature of their requirements and results. This naturally suggests that by properly

* Joint appointment with the Research Institute for Informatics, Bucharest, Romania

integrating them, one could obtain a synergistic effect in which different strategies mutually support each other, and compensate for each other's weaknesses. As a result, such a system may be applicable to a wide spectrum of problems. This is the motivating idea behind the growing interest in developing multistrategy learning systems (Birnbaum and Collins, 1991).

The aim of our research is to develop a theoretical framework and a methodology for multistrategy task-adaptive learning (MTL) systems that would integrate a whole range of learning strategies (Michalski, 1990; Tecuci and Kodratoff, 1990; Tecuci and Michalski, 1991). Given a learning problem, the system should be able to determine which strategy or a combination of strategies would be most effective to apply. It would also attempt to derive some "useful" information from every type of input, regardless as to whether it is a new fact, a general piece of knowledge, knowledge similar to a part of the learner's prior knowledge, or even a fact already known.

This paper describes an MTL method that generalizes and extends the ideas implemented in the DISCIPLE system (Tecuci, 1988; Tecuci and Kodratoff, 1990) by incorporating new reasoning and generalization methods, and by extending the types of knowledge that a learner is able to acquire. The key idea of the learning method is that the system tries to "understand" the input in terms of its current knowledge, and uses this understanding to improve the knowledge.

2 The learning task

The *learning task* of a learning system is defined as a triple consisting of the learner's goal(s), its current knowledge base (KB), and the input information.

The method proposed in this paper assumes that the learner starts with an initial knowledge base, which may be incomplete and/or partially incorrect. This knowledge base may include a variety of knowledge types: facts, concept examples, different kinds of rules (e.g. decision rules, causal relationships, determinations, mutual dependencies or implications), structural object descriptions, concept and relationships hierarchies, as well as representations of the learner's capabilities and actions in the world.

The input can be a fact, an example of some concept or relationship, a specific solution of some problem, or several such knowledge pieces.

It is also assumed that the learning goal is to continually extend, update and improve the learner's knowledge. Such a general goal subsumes different more specific goals, such as to improve efficiency of the knowledge base by reorganizing some parts of it, to acquire new general pieces of knowledge, to correct the knowledge base in view of new experiences, etc.

As will be shown in the following, the proposed MTL method "adapts" to a given learning task by applying a learning strategy, or a combination of strategies, that are most suitable for it.

3 Understanding the input

Whenever the system receives an input, it tries to "understand" it in the context of its KB (Tecuci and Kodratoff, 1990). This means that the system tries to demonstrate that the input is a plausible consequence of the knowledge it already has, or that it represents new knowledge. The proposed method performs such a process by building a *justification tree*. To describe the method, let us assume that the input is an example of some relationship, $P_n(x,y)$, where P_n is the name of the relationship, between objects x and y that satisfy it:

$$P_1(a,f) \ \& \ P_2(g,a) \ \& \ P_3(b) \ \& \ P_4(h) \ \&...\& \ P_i(b,e) :: > P_n(a,b) \quad (1)$$

$P_1,...,P_i$ are predicates, 'a', 'b', 'f'... , 'e' are object names or object properties, and : : > denotes a *target assignment operator,* which assigns "truth" to the predicate on the RHS, if the expression on the LHS is true. For illustration, below is such an example of the relationship "grows(x, y)":

rainfall(Vietnam, heavy) & climate(Vietnam, subtropical) & soil(Vietnam, red-soil) & location(Vietnam, SE-Asia) : : > grows(Vietnam, rice) (2)

To "understand" the example stated generally in eq.(1), the system attempts to demonstrate that $P_n(a,b)$ is a plausible consequence of $P_1(a,f)$ & $P_2(g,a)$ & $P_3(b)$ & $P_4(h)$ &...& $P_i(b,e)$, in the context of the learner's background knowledge (BK). To make such a demonstration, the system builds a plausible justification tree (Figure 1). The concept of plausibility is used here in the sense described in (Collins and Michalski, 1989). This is different from the concept of plausibility used by (DeJong, 1989), which is based on qualitative reasoning.

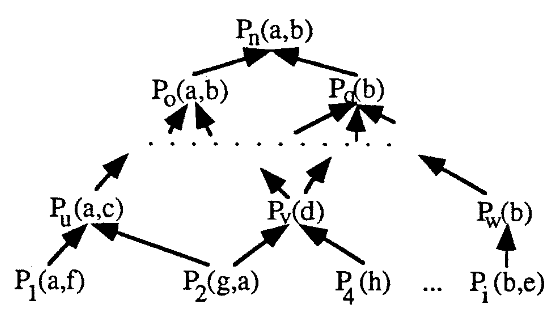

Figure 1. A plausible justification tree.

The root of the tree in Figure 1 is the RHS of eq.(1). The leaves are the predicates on the LHS of eq.(1), and the intermediate nodes are intermediate predicates generated during the "understanding" process. The branches connected to any given node link this node with predicates, the conjunction of which *plausibly implies* the predicate at the node, according to the learner's BK. The notion "plausibly implies" means that the target (parent node) can be inferred from the premises (children nodes) by deduction, induction or analogy, using the learner's BK. The branches together with the nodes they link represent individual inference steps. To indicate such inference steps in a general way, we use the symbol "-->", called the *inference operator.*

For example, the inference step

$$P_1(a,f) \ \& \ P_2(g,a) \dashrightarrow P_u(a,c) \tag{3}$$

(see Figure 1) may be a result of a deductive step based on the implication residing in BK:

$$\forall s \ \forall t \ \forall z \ \forall p \ (P_1(s,t) \ \& \ P_2(z,s) \Longrightarrow P_u(s,p)) \tag{4}$$

Other inference steps may employ some form of plausible inference. For example, the step

$$P_2(g,a) \ \& \ P_4(h) \dashrightarrow P_v(d) \tag{5}$$

may be a result of analogical inference. To illustrate such a step, suppose that BK contains the implication (6). Suppose further that the system determined that g, a, h and d are similar to g', a', h' and d', respectively, in the context of the properties considered relevant to the learning problem (Collins and Michalski, 1989). By analogy, the system concludes that from $P_2(g,a)$ & $P_4(h)$ one can plausibly infer $P_v(d)$, and hence (5).

$$P_2(g',a') \ \& \ P_4(h') \Longrightarrow P_v(d') \tag{6}$$

The above is a very simple form of analogy. In general, the method is intended to incorporate different forms of analogy, in which the analogical transfer is based on different kinds of similarities, such as similarities among relations, causes, and meta-relations.

Another type of plausible reasoning is induction. As shown in (Michalski, 1990), induction can have different forms: abduction, empirical generalization, or constructive induction (the latter one may incorporate the previous two, as well as deduction). To illustrate an abductive inference step, suppose that BK contains a causal relationship

$$\forall r \ (P_w(r) \Longrightarrow P_i(r,e)) \tag{7}$$

By matching the predicate Pi(b,e) in eq.(1) with the RHS of (7), the system hypothesizes through abduction $P_w(b)$. Hence, we have a plausible inference step:

$$P_i(b,e) \dashrightarrow P_w(b) \tag{8}$$

An inference step could also be done through a combination of empirical generalization and deduction. To illustrate this, let us suppose that the system failed to prove that "$P_n(a,b)$" derives deductively, analogically or abductively, from other predicates. In such a case, it will look for examples in which P_n is true. If such examples can be found in the BK, then the system tries to inductively generalize them in order to learn a rule that would solve the problem at hand. To illustrate this, let us suppose that the system found in its BK the examples:

$$P_o(c, f) \;\&\; P_q(c) \;\&\; P_j(m) \;::>\; P_n(c, f) \qquad (9)$$
$$P_o(d, g) \;\&\; P_q(d) \;\&\; P_k(n) \;::>\; P_n(d, g)$$

These examples can be empirically generalized to the rule

$$\forall x \forall y \, (P_o(x,y) \;\&\; P_q(y) \Longrightarrow P_n(x,y)) \qquad (10)$$

Rule (10) is then used to produce the plausible inference step $P_o(a, b)$ & $P_q(b)$ --> $P_n(a, b)$. In a more complex case, available examples may not be so easily generalizable to (10), and the system may have to use constructive induction.

For instance, a plausible justification tree for the example in (2) is the following one (Tecuci and Michalski, 1991):

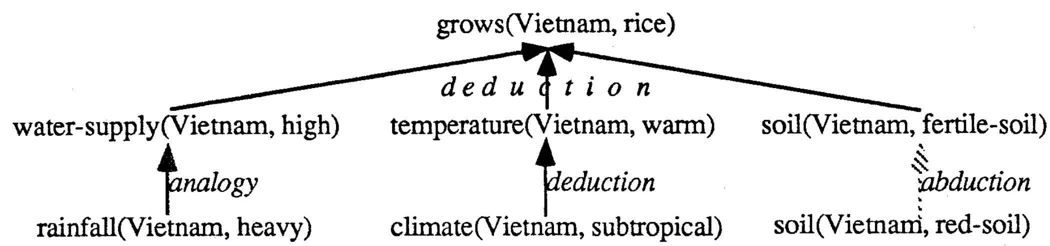

Figure 2. A plausible justification of the example in (2).

The above shows that an inference step in a justification tree may be a result of any type of inference - deductive, analogical or inductive (where the latter may be in the form of empirical generalization, abduction or constructive induction). The type of inference applied is a result of the system's attempt to demonstrate how the input relates to the BK.

A natural question is which type of inference is actually used, when more than one inference type applies at a given step, and they produce different conclusions. This is, in fact, a frequent situation in human reasoning, in which different "lines of reasoning" may produce different results. (Collins and Michalski, 1989) argue that people estimate the "strength" of different lines of reasoning, and make their conclusion on the basis of this evaluation. If the lines lead to the same conclusion, they have a strong belief in the result. If the lines lead to different conclusions, and the associated "strengths" are roughly similar, people restrain from making any decisive conclusion.

We have not yet conclusively investigated this issue. In the proposed method the system follows the following control strategy: first, it tries to justify a given predicate by deduction. If this attempt fails, then it tries to justify it by a similarity to a specific past case, and then by an analogy to some rule in BK. If these strategies fail, the system tries an abduction from relevant rules. With the lowest preference, it tries to collect facts or examples which can be generalized to a rule that can be used to justify the given predicate. Once such a rule is generated, the justification can be done using the rule as any other rule, i.e., deductively, abductively or analogically.

The understanding process proceeds in the same way when the input is a new fact, a new relationship or a specific solution of some problem. Indeed, let us suppose that the input is just the relationship $P_n(a,b)$, which is interpreted as true knowledge. In such a case, the system is building the plausible justification tree from Figure 1 that shows that the relationship $P_n(a,b)$ is a plausible consequence of the following facts and relationships from the knowledge base: $P_1(a,f)$ & $P_2(g,a)$ & $P_4(h)$ &...& $P_i(b,e)$.

If the input is a specific solution S to some problem P, then the plausible justification tree represents a demonstration that S solves P.

4 Generalizing the understanding

The next step of the proposed MTL method consists of generalizing the plausible justification tree as much as allowed by the knowledge used to derive it in the first place. The technique used is similar to that developed by (Mooney and Benett, 1986) for explanation-based learning (DeJong and Mooney, 1986; Mitchell et al., 1986). First, each inference step from the plausible justification tree is generalized locally and then all these generalizations are globally unified, giving a generalized justification tree. The difference is that our method employs several techniques for the generalization of the individual inference steps, which depend on the type of inference and the system's background knowledge.

To illustrate this process, take, for instance, the inference step (3), i.e., $P_1(a,f)$ & $P_2(g,a)$ --> $P_u(a,c)$. This inference step can be locally generalized into the rule (4) that allowed this step, i.e., $\forall s\ \forall t\ \forall z\ \forall p\ (P_1(s,t)\ \&\ P_2(z,s) ==> P_u(s,p))$. The branches of the tree in Figure 1, corresponding to the original inference step (3), are then replaced by the appropriate components of this rule. This is a deductive generalization step.

Let us now consider the branch corresponding to the inference step (5), i.e., $P_2(g,a)$ & $P_4(h)$ --> $P_v(d)$. This step is based on analogy with implication (6), i.e., $P_2(g',a')$ & $P_4(h')$ ==> $P_v(d')$, which was a part of BK. The generalization step is based on the idea that a similarity of an entity to a given entity generates an equivalence class of all entities similar to the given entity. Following this idea, the system generates a conjunctive generalization that covers all the inference steps that could be derived by analogy with (6):

$$P_2(g^*,a^*) \ \& \ P_4(h^*) ==> P_v(d^*) \qquad (11)$$

where g*, a*, h* and d* represent classes that contain g and g', a and a', h and h', d and d', respectively. A simplified version of this procedure is to replace the inference step (5) with the most specific conjunctive generalization of (5) and (6).

As mentioned previously, there are different forms of analogy, in which the analogical transfer is based on different kinds of similarities. Our claim is that each analogical inference should be generalized according to a specific procedure, which takes into account the kind of knowledge used in analogy (determinations, mutual dependencies, etc.).

The abductive step $P_i(b,e)$ --> $P_w(b)$ is replaced by $P_i(r,e)$ --> $P_w(r)$, according to the general rule (7), i.e. $\forall r$ ($P_w(r)$ ==> $P_i(r, e)$). This generalization is justified because any system making such an abduction could also make the abduction $P_i(r,e)$ --> $P_w(r)$, for any r.

Finally, the inference step $P_o(a,b)$ & $P_q(b)$ --> $P_n(a,b)$ was done by applying the rule (10), i.e. $\forall x \forall y(P_o(x,y)\&P_q(y)$ ==> $P_n(x,y))$, which was obtained by empirical generalization. Then the corresponding branch is replaced by rule (10).

As mentioned above, after each inference step is locally generalized, the system globally unifies the patterns. Such a process produces the generalization of the plausible justification tree in Figure 1 (see Figure 3).

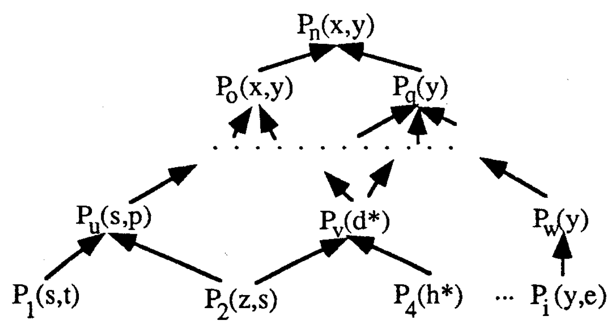

Figure 3. A generalization of justification tree from Figure 1.

To give a specific example, a generalization of the plausible justification tree in Figure 2 is indicated in Figure 4 (Tecuci and Michalski, 1991).

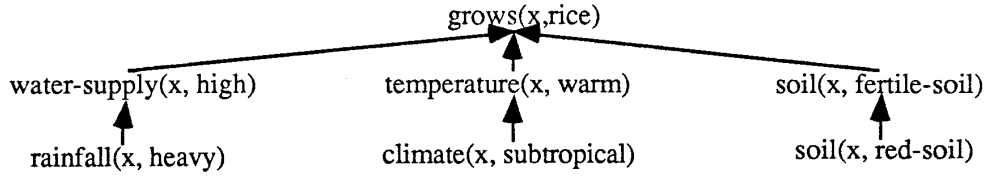

Figure 4. A generalization of the plausible justification tree from Figure 2.

5 The learned knowledge

A distinctive feature of the presented MTL method is the great diversity of the types of knowledge it can learn.

During the "understanding" of the input, the system may learn (by analogy, abduction, or generalization) new pieces of knowledge (facts, relationships, specific inferences, or even general rules) that help building of the plausible justification:

$P_2(g,a)$ & $P_4(h)$ --> $P_v(d)$	(learned by analogy)
$P_i(b, e)$ --> $P_w(b)$	(learned by abduction)
$\forall x \forall y\ (P_o(x,y)\ \&\ P_q(y) ==> P_n(x,y))$	(learned by empirical generalization)

During the generalization of the plausible justification, some of the previously learned knowledge may also be generalized:

$P_2(g^*,a^*)$ & $P_4(h^*)$ ==> $P_v(d^*)$	(learned by empirical generalization)

Depending on its goals, the system may also extract several pieces of knowledge from the generalized justification tree. One is a definition of the relationship $P_n(x,y)$, expressed in terms of predicates present in the input example (an "operational" definition):

$$P_1(x, t)\ \&\ P_2(z, x)\ \&\ P_4(h^*)\ \&...\&\ P_i(y, e)\ ::>\ P_n(x, y) \qquad (12)$$

Another component is the most abstract characterization of the relationship $P_n(x,y)$, based on the top part of the justification tree:

$$P_o(x,y)\ \&\ P_q(y) ==> P_n(x,y) \qquad (13)$$

Other components are various abstractions of the input example. For instance, one abstraction is obtained by instantiating the variables in the above abstract characterization (eq.(13)), to specific arguments in the original example:

$$P_o(a,b)\ \&\ P_q(b)\ ::>\ P_n(a,b) \qquad (14)$$

Other abstractions would correspond to lower levels of the generalized justification tree.

6 Learning from multiple examples

In the previous sections we have outlined multistrategy task-adaptive learning from one positive example. However, the presented method can be extended to learn from several (positive and negative) examples, as will be briefly presented in the following.

From the first positive example the system is building the plausible justification trees in Figures 1 and 3.

For each new positive example E_i, the system will generalize the plausible justification tree in Figure 3 so as to cover a plausible justification for E_i. Also, some of the knowledge pieces from BK may be generalized so as to cover inferences from the plausible justification of E_i.

For each new negative example N_j, the system will specialize the general plausible justification tree and the KB so as to no longer cover any plausible justification for N_j. The technique employed is similar to the one used in the DISCIPLE system (Tecuci, 1988; Tecuci and Kodratoff, 1990). First, the system creates the instance of the general plausible justification tree corresponding to the negative example N_j. This is a wrong justification tree because it "proves" that N_j is an example of the relationship (concept) to be learned. Next, the system hypothesizes which inference step in this tree is the false one, and removes it from the system's knowledge, by specializing an inference in the general justification tree, and some knowledge pieces from the BK. This process may be guided by a minimum-cost-change heuristic which determines the inference, the removal of which produces the least changes in system's knowledge. The elementary criteria to be used in estimating the cost of the change may include the strength of the inference, the loss of coverage through specialization, and the increase in complexity. These criteria may be combined into one general measure using the lexicographic evaluation functional or LEF (Michalski, 1983).

7 Conclusion

We have shown how input "understanding" can form a natural basis for a multistrategy task-adaptive learning method. The major advantage of this method is that it enables the system to learn in situations in which single-strategy learning methods, or even previous integrated learning methods were insufficient. Also, it has a high degree of generality, being able to cope with a great variety of input information and knowledge pieces in the background knowledge.

There are many possible extensions of the proposed MTL method that we intend to address in our future research. For instance, learning from multiple examples has to be refined. Also, the control strategy may be improved by following the idea of the different "lines of reasoning" mentioned in section 3. We also plan to consider errors in the input. In such a case, the system has to decide if the cause of a contradiction between the input and the BK is the incorrectness of the input, or that of the BK, or both, and to take appropriate corrective actions. Another possible extension regards the integration into the MTL method of additional learning strategies like conceptual clustering and reinforcement learning. We also plan to extend and use the method for extracting knowledge from a human expert.

Acknowledgments

This research was done in the Artificial Intelligence Center of George Mason University. The activities of the Center are supported in part by the Office of Naval Research under grant No. N00014-88-K-0226, No. N00014-88-K-0397, and No. N00014-91-J-1351 and in part by the Defense Advanced Research Projects Agency under grant, administered by the Office of Naval Research, No. N00014-87-K-0874, and N00014-91-J-1854.

References

Birnbaum, L., and Collins, G. (eds.), *Machine Learning: Proceedings of the Eighth International Workshop,* San Mateo, CA, Morgan Kaufmann, 1991.

Collins, A., and Michalski, R.S., The Logic of Plausible Reasoning: A Core Theory, *Cognitive Science*, Vol. 13, No. 1, pp.1-49, 1989.

Davies T.R., and Russell S.J., A Logical Approach to Reasoning by Analogy, *Proceedings of IJCAI-87*, pp. 264-270, Milan, Italy, 1987.

DeJong, G. and Mooney, R., Explanation-Based Learning: An Alternative View, *Machine Learning Journal*, vol 2, 1986.

DeJong, G., Explanation-based Learning with Plausible Inferencing, *Proceedings of the Fourth European Working Session on Learning* (EWSL89), Montpellier, France, 1989.

Michalski, R. S., Theory and Methodology of Inductive Learning, *Machine Learning: An Artificial Intelligence Approach*, R.S.Michalski, J.G.Carbonell, T.M.Mitchell (Eds.), Tioga Publishing Co., 1983.

Michalski, R.S., Toward a Unified Theory of Learning: Multistrategy Task-Adaptive Learning, in *Reports of Machine Learning and Inference*, MLI 90-1, Center for Artificial Intelligence, George Mason University, January 1990.

Mitchell, T.M., Keller, T., and Kedar-Cabelli, S., Explanation-Based Generalization: A Unifying View, *Machine Learning Journal*, Vol. 1, January 1986.

Mooney R., Bennet S., A Domain Independent Explanation Based Generalizer, in *Proceedings AAAI-86*, Philadelphia, 1986, pp.551-555.

Tecuci G., DISCIPLE: A Theory, Methodology, and System for Learning Expert Knowledge, *Ph.D. Thesis*, University of Paris-South, 1988.

Tecuci G. and Kodratoff Y., Apprenticeship Learning in Imperfect Theory Domains, in *Machine Learning: An Artificial Intelligence Approach vol. III*, Morgan Kaufmann, 1990.

Tecuci G., and Michalski R.S., A Method for Multistrategy Task-adaptive Learning Based on Plausible Justifications, in Birnbaum L., and Collins G. (eds), *Machine Learning: Proceedings of the Eighth International Workshop,* Chicago, June 1991, Morgan Kaufmann, 1991.