TYPES OF EXPLANATION AND THEIR ROLE IN CONSTRUCTIVE CLOSED-LOOP LEARNING

by

H. Ko
R. S. Michalski

Reports of Machine Learning and Inference Laboratory
George Mason University, 1988.

TYPES OF EXPLANATION and THEIR ROLE IN CONSTRUCTIVE CLOSED-LOOP LEARNING

H. Ko
Department of Computer Science
University of Illinois
1304 W. Springfield
Urbana, IL 61801
Tel: 217-333-1376

and

R. S. Michalski
Artificial Intelligence Center
George Mason University
4400 University Drive
Fairfax, VA 22030

ABSTRACT

The paper discusses a distinction between the knowledge used by an explanation process (*explanation knowledge*) and the structure built as a result of this process (*explanation structure*). An explanation of any data or phenomenon is viewed as a composition of these two components. Two types of explanations are discussed: *deductive* and *inductive*. Based on these ideas, *empirical*, *analytical* and then *constructive* learning methods are compared.

Constructive closed-loop learning (or, briefly, constructive learning) is an integrated form of learning, which uses deductive inference when the observation is logically implied by the background knowledge, and employs inductive inference when a new piece of knowledge is necessary to establish the explanation structure. An example of constructive learning as applied to a problem in robot assembly is presented and analyzed in detail.

1. Introduction

An explanation of some phenomenon involves finding knowledge that logically entails the phenomenon, and creating a structure that demonstrates that this knowledge indeed entails the phenomenon. This distinction is useful for gaining a uniform perspective of various learning paradigms, and leads us a formulation of a new learning paradigm unifying inductive and deductive learning (Michalski and Ko, 1988; Michalski and Watanabe, 1988).

To illustrate this distinction, suppose that a person did not come to a meeting with a friend. Suppose that this person later told his friend that he did not come to the meeting because his mother got sick. This statement, called *explanation knowledge*, explains the person's behavior to his friend, if the friend has the background knowledge that a crisis in a family typically overrides other commitments, and that a mother's sickness is a family crisis. The explicit demonstration that the explanation knowledge indeed explains the behavior (which in our example case is a simple two step application of modus ponens) is called *explanation structure*. The explanation structure is thus a proof that the explanation knowledge together with background knowledge logically implies the behavior, or, in general, any phenomenon that is being explained.

In analytic learning, such as explanation-based learning, without the knowledge of the person's mother illness, no explanation can be constructed, and the system stops. People in such cases, however, usually construct some hypothesis to explain the faced phenomenon, in this case, the absence.at a meeting. The friend, for example, may hypothesize that something serious might have happened to the person . This leap of faith is a form of constructive inductive inference (Michalski, 1983). It is done by employing the background knowledge that if a person, who normally comes to meetings, did not come, then it is likely that something serious is the reason.

To hadle such problems, *constructive closed-loop learning* (or, briefly, *constructive learning;* Michalski and Watanabe, 1988) uses deductive inference when the observation is logically implied by background knowledge, but resorts to inductive inference when an additional (or modified) knowledge is necessary to establish the needed explanation structure. Thus, such a system can learn when its initial knowledge is inadequate. Many practical learning situations start with such an inadequate knowledge. An important example of such a situation is learning by an autonomous robot exploring a partially known environment., or a robot assembling a device without a complete control knowledge of the procedure.

Next section uses the above ideas to distinguish between two types of explanation, *deductive* and *inductive*. Then, based on this distinction, different learning approaches are discussed. Constructive learning is characterized in terms of these ideas, and illustrated by a detailed analysis of an example from the area of robot assembly.

2. Types of Explanation

To explain some observation to an agent means to construct a knowledge structure that conceptually relates the agent's background knowledge (BK) with the observation statement (OS). Formally, the constructed knowledge structure, called *explanation structure* (ES), must demonstrate that BK logically entails OS, which we write:

$$\text{BK} \mid\!> \text{OS} \qquad (1)$$

In other words, OS must be a logical consequence of BK. In many situations, however, the background knowledge (BK) may not be adequate to establish (1). BK may be inadequate because it may be insufficient, intractable (too complex), or inconsistent with the observations. In all such cases, BK has to be modified or enhanced by additional knowledge, called *explanation knowledge* (EK). The explanation knowledge may be given to us from another source, e.g., teacher or environment, or may have to be hypothesized. In these cases, the explanation structure (ES) is a proof that

$$\text{BK} \ \& \ \text{EK} \mid\!> \text{OS} \qquad (2)$$

Constructing an explanation knowledge may require changing (updating, correcting, etc.) BK into some modified BK*. Thus, in general, (2) is in the form

$$\text{BK*} \ \& \ \text{EK} \mid\!> \text{OS} \qquad (3)$$

Based on these considerations , we can say that an explanation of an observation consists of two components:

EK - explanation knowledge

ES - an explanation structure that demonstrates that explanation knowledge together with background knowledge logically entails observation .

Using this conceptual framework, we can unify different methods of learning. For example, in explanation-based learning, EK is null, and one seeks the explanation structure, ES, that shows (1). In empirical learning, BK is small and inadequate for explaning OS, so one needs to apply inductive learning to hypothesize an explanation knowledge EK, based on the input data. In constructive induction BK may be substantial, but still inadequate for deductively explaning the observation. There may be many different situations between the two extremes, the purely empirical and purely analytic learning. When BK is inadequate, and one needs either to create EK so that (2) holds, and/or modify BK so that (3) holds. Constructive closed-loop learning is an attempt to develop a system that can handle all such situations.

The above leads us to the distinction between two types of explanation:

- a *deductive explanation,* which consists merely of the explanation structure demonstrating that the observation, OS, is a logical consequence of what the system already knows (BK[1]), i.e., that BK |> OS. The explanation knowledge (EK) is null .

- an *inductive explanation,* which consists of an explanation knowledge (EK), which is inductively hypothesized, and the explanation structure, which demonstrates that EK together with (possibly modified) background knowledge, BK*, it implies the observation: (OS), i.e., BK* & EK |> OS .

On the basis of these concepts, we will analyze in more detail the empirical, analytical and constructive learning.

3. Empirical Learning

Empirical learning presupposes little background knowledge relevant to the task at hand, so that the main concern is to hypothesize a concept or rule primarily on the basis of the observational data supplied to the system [Michalski, 1983, 1987]. Since there is usually a plethora of possible hypotheses that could explain an observation, the main problem is to find the most plausible, or generally, the most preferred explanation . Thus, the main inference scheme of these systems is inductive.

[1] In explanation based learning literature, BK^* is typically called domain knowledge. In general, BK contains domain-specific knowledge, domain independent knowledge, and metaknowledge (rules of inference, constraints, etc.).

The empirical learning task is described as follows:

Given:

- Observational statements (OS) about an object, phenomenon, or a process.
- Background knowledge (BK) which includes domain concepts, the preference criterion for choosing among competing hypotheses, and inductive rules of inference.

Determine:

- Explanation knowledge (EK) that, if true, logically entails the observation and is most plausible, or, in general, most desirable among all other such hypotheses according to a given preference criterion.

Explanation structure (ES) in most EIL systems involves subsumption relationships between EK and OS. Thus, they use a matching procedure rather than a full deductive decision procedure to establish ES. For example, in SPARC/G [Michalski & Ko & Chen 1987], the observational statements are a sequence of snapshots of some unknown process. The background knowledge includes attributes used to describe the snapshots, associated types and structures of the value sets of the attributes, a rule preference criterion, and generalization rules. In addition, it includes description models that constrain the form of plausible explanation knowledge. The system determines rules (EK) for each description model that would qualitatively predict the future continuation of the unknown process.

Analytic Learning

Most known forms of analytical learning are explanation-based generalization [Mitchell & Keller & Kedar-Cabelli 1986] and explanation-based learning [DeJong & Mooney 1986]. In this approach the system attempts to show that the background knowledge the observer possesses accounts for the observation. A successful explanation enable the system to formulate a more efficient or operational rule for accounting for the observation. The basic inference scheme used in these systems is deductive. Analytic learning can be described as follows:

Given:

- Observational statements (OS) about some objects, phenomena, or processes

- Background knowledge (BK) which contains general and domain-specific concepts for interpreting the observations, as well as relevant inference rules.

Determine:

- A reformulation of the background knowledge that logically entails the observation and is more effective and/or efficient. then the prior knowledge.

The explanation structure (ES) in these systems is either a proof tree generated by a theorem prover, a trace of the Horn clauses in Prolog, or some other equivalent form. For example, in ARMS system (Segre 1987), the observational statements include a joint relationship between two components of an assembly (goal statement) and a sequence of actions performed by a teacher that achieves the joint relationship successfully. The background knowledge includes general plans for achieving simpler joint relationships. The explanation process identifies general plans in the background knowledge that participated in the teacher's actions and establishes a sequence of plan instances (ES). Then, the learner determines a new general plan for achieving the joint relationship in the goal statement from ES, a reformulation of background knowledge.

Constructive Closed-loop Learning

Empirical learning, described above, is one form of inductive learning. Another form is constructive induction (Michalski 1983). Constructive induction incorporates domain specific knowledge and a deductive reasoning mechanism into an empirical learning system. The learning system uses this knowledge to construct new descriptors and concepts not present in the input data, in order to generate more adequate inductive hypotheses. Recently, the concept of constructive induction was generalized to *constructive closed-loop learning*, or, briefly, *constructive learning* (Michalski, 1987; Michalski and Watanabe, 1988).

Constructive learning integrates inductive learning (empirical and constructive) with deductive (analytical) learning, and includes also the ability to determine the task relevant knowledge in the knowledge base, and to evaluate the constructed knowledge in order to decide if it is to be stored in the knowledge base for future use. A CCL system is able to use full deductive decision procedure to establish explanation structure, as well as to synthesize plausible explanation knowledge, like empirical learning systems. Such an integration is needed for a number of applications, for example, for implementing earning in intelligent robots.

A prototype CCL system is currently being developed for the Intelligent Explorer Project (IEX), conducted jointly by the George Mason University and University of Illinois.. The goal of the project is to develop an autonomous robot capable of learning, reasoning and planning in a partially known environment. The system distinguishes knowledge sources that are modifiable (hypotheses and/or beliefs) from principles and definitions that are irrefutable. .All the inference steps taken by the system are recorded, using assumption-based truth maintenance system (de Kleer, 1986).

If any contradictions are detected, the system identifies the knowledge sources which participated in the justification structure and searches for the blame from the most modifiable knowledge. Various modification strategies are applied to avoid similar type of contradictions. The revised knowledge participates in the explanation process and is subject to further scrutiny for their validity. Due to the space limitation, we restrict ourselves to some central aspects of constructive learning, and explain how the system works by going through an example in the area of automatic assembly.

3. EXAMPLE: Learning Assembly Sequences

Suppose a robot can only carry out one assembly task between two parts at a time and the assembly operation is successful as long as there is a collision free trajectory for one part to mate with the other. So, the main task of the robot planner is to order the tasks for achieving individual spatial relationships prescribed in the final assembly so that the plan is executed without any collision. The problem for the *robot planning module* is defined:

Given:

- Parts of an assembly and their geometric features.
- Initial configuration of the parts.
- The spatial relationships between parts that exist in the final assembly and their associated assembly tasks.
- Task ordering knowledge.

Determine:

- Temporal orderings of the individual assembly tasks.

Once the system learned the temporal orderings, individual assembly instructions are then executed. Each assembly operation is monitored to determine if the trajectory planner fails to find a collision free trajectory. If it fails, the monitor reports all the objects that are blocking the path of the object being moved during an assembly operation. The function of the *execution monitor* is defined below:

Given:

- Individual assembly tasks
- Temporal orderings determined by the robot planner

Determine:

- All the objects that blocked the trajectory when an assembly operation failed.

The learning task is to recognize these failures and rectify the situation by modifying the task ordering knowledge. The function of the *learner* is defined as follows:

Given:

- Parts that are being assembled.
- Spatial relationships between parts.
- Task orderings of individual assembly tasks that were generated by the robot `s planner.
- All the objects that were in the way during the failed assembly operation determined by the execution monitor.

Determine:

- A modification of the task ordering knowledge that is responsible for the failure so that similar type of failure will not happen in the future.

Suppose the robot is to assemble the head portion of a bell as shown in figure 1 and the planner first generates a plan: assemble ring to the pin and then, assemble bell-head to the pin. The second assembly instruction fails because both the bell-head and the ring are in the way. In a sense, this is part of the explanation knowledge from the environment.

However, it is not sufficient. We would like to create task ordering knowledge that would predict this failure in the future. We have:

Observational Statements

1. Bell-head and ring are being assembled to the pin.
2. The spatial relationships, mating conditions, between bellhead, ring, and pin: e.g., Aligned(pin, bell-head), Against(pin, bell-head), Aligned(pin, ring), Against(ring, bellhead).
3. Bell head is above the ring with respect to z-axis of the world inertial reference frame in the final assembly.
4. Bell head is below the ring with respect to z-axis of the base frame of the pin.
5. Task ordering: Assemble(pin, ring) < Assemble(pin, bell head)[2]
6. The execution monitor notices both pin and ring are in the way and a simple interpretation module asserts: not (Assemble(pin, ring) < Assemble(pin, bell head)).

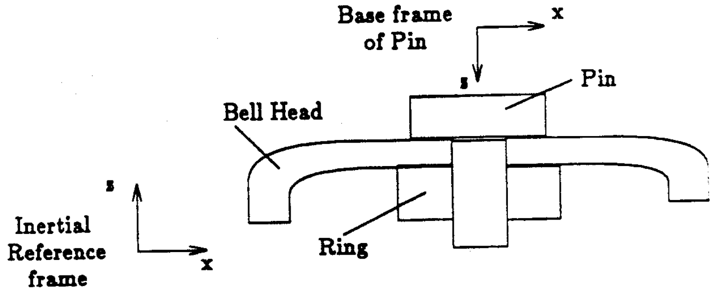

Figure 1. Assembly of Bell Head

Let us consider different scenarios for constructing an explanation structure to answer why the assembly plan failed. First, assume that only domain specific descriptors about objects between objects in the observation are known.

[2] Task1 < Task2 means Task1 should be carried out before Task2. So, in this case, the ring is assembled to the pin before the bell head.

Background Knowledge:

7. Geometric desriptions of object types, pin, bell head, and ring including the base frame of the pin[3]

8. Spatial reasoning engine that compute relative spatial relationships. kinematic degrees of freedom, and more[4] [Ko ,1987].

An empirical learning system would postulate a surface explanation knowledge that the plan failed because, for example:

- the bell head is above the ring with respect to z-axis of the inertial reference frame,

or

- the bell head is below the ring with respect to z-axis of the base frame of the pin.

There are many more possible EK's. However, there is little guidance as to which is more plausible than the other. So, the system resorts to multiple explanations of the same phenomenon. In the second scenario, the observer may possess more extensive knowledge of the environment.

Background Knowledge:

7. Geometric desriptions of object types, pin, bell head, and ring.

8. Spatial reasoning engine.

9. When parts are assembled to a shaft, assemble them bottom up with respect to z-axis of the base frame of the shaft.

10. A pin is a kind of shaft.

Here, the explanation structure is constructed using a deductive decision procedure.

Explanation Structure:

11. The task ordering, assemble(pin,ring) < assemble(pin,bell head), failed because 1, 4, 8 and 9.

[3] Each part is associated with its base frame and all the geometric features as defined with respect to the base frame.

[4] Each part is associated with its base frame and all the geometric features are defined with respect to the base frame.

Although this is only one step deduction, in general, multiple inference steps are required to establish the explanation structure. A temporal reasoning module usually requires multiple inference steps to detect temporal constraint violation but because of the space limitation, only one step violation is shown. In analytic learning, a reformulation of initial knowledge is sought that would make it more effective and/or efficient:

- When parts are being assembled to a pin, assemble them from bottom up along the z-axis of the pin.

Finally, the third scenario involves both capabilities, analytical and empirical learning. After the first example for bell head assembly, the learner postulates competing hypotheses using empirical learning capabilities:

1. When partI is above part2 with respect to z-axis of the world inertial reference frame in the final assembly, assemble part1 first.

2. When part1 is below part2 with respect to z-axis of the base frame of part3, assemble part1 first.

Then, the robot is presented with a new problem of mounting a bracket as shown in Figure 2.

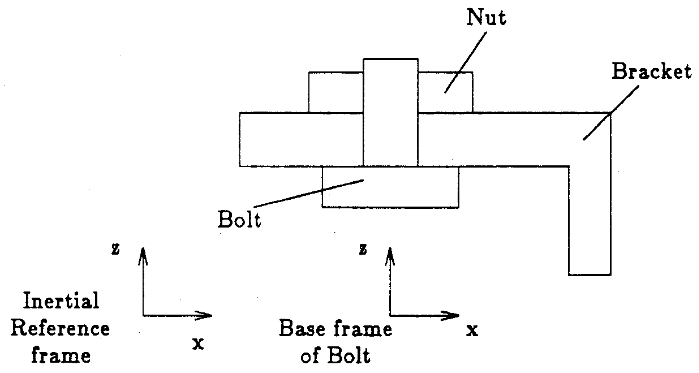

Figure 2. Bracket Mounting Assembly

Using the first explanation as a task ordering knowledge, nut is assembled to the bolt before the bracket. This fails. On the other hand, the second hypothesis correctly predicts that the bracket should be assembled before the nut. So, the hypothesis 2 is more plausible than hypothesis 1.

This example have shown how a constructive learning sytem can handle a learning problem that neither empirical nor analytic learning system system could.

Summary

We have introduced a distinction between a deductive and inductive explanation. This distinction was then used to characterize empirical and analytic learning systems. We have then discussed constructive learning, which integrates empirical and analytic learning. A constructive learning system was described by a detailed analysis of an example problem and solution in the area of automated assembly.

The paper dealt only with problems of detemining and emplying different types of explanation in constructive learning. Other aspect of such learning, such as determination of relevant knowledge, and an evaluation of generated knowledge was outside of the scope of this paper. The presented work indicates that constructive learning, which integrates different learning paradigms within a single conceptual framework, is an important new research direction for machine learing.

Acknowledgements

The authors thank the members of the Intelligent Systems Group, and the Intelligent Explorer Group. for discussion and criticism of ideas presented here

This research was supported in part by the Office of Naval Research under grant No. N00014-82-K-0186, and in part by the Defence Advanced Project Agency under the grant administered by the Office of Naval Resaerch No. N00014-K-85-0878.

References

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

de Kleer, J., An Assumption-Based Truth Maintenance System, *Artificial Intelligence,* vol. 28, no. 1, 1986.

Ko, H., Reasoning about Kinematics from Mating Conditions, submitted for publication to *IEEE Journal of Robotics and Automation,* October, 1987

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., Concept Learning, *Encyclopedia of Artificial Intelligence,* E. Shapiro (ed.), John Wiley & Sons, January, 1987.

Michalski, R. S., Ko, H. and Chen, K, Qualitative Prediction: The SPARC/G Methodology for Describing and Predicting Discrete Processes, in Expert Systems, P. Dufour and A. Van Lamsweerde (eds), Academic Press Incorporated, pp 125-158, 1987.

Michalski, R.S.and Ko, H.,. On the nature of explanation or Why did the wine bottle shatter, A paper accepted for the AAAI Workshop on Explanation-based Learning, Stanford University, March 1988.

Michalski, R. S., Watanabe, L., Constructive Closed-loop Learning: Fundamental Ideas and Examples, a paper submitted for presentation at the Fifth International Conference on Machine Learning, Ann Arbor, Michigan, June 12-15, 1988..

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

Segre, A. M., Explanation-Based Learning of Generalized Robot Assembly Plans, PhD thesis, UILU-ENG-87-2208, Coordinated Science Laboratory, Univeristy of Illinois, January 1987.