An Integrated Discovery System

Bernd Nordhausen
Pat Langley

Irvine Computational Intelligence Project
Department of Information & Computer Science
University of California, Irvine CA 92717
phone: (714) 856-0935

bernd@cip.uci.edu
langley@cip.uci.edu

Abstract

Previous discovery systems have successfully rediscovered known scientific laws. However, they have addressed only limited parts of the empirical discovery task. In this paper we introduce an integrated discovery system (IDS) which addresses many issues of empirical discovery. IDS operates in three stages: taxonomy formation, qualitative discovery, and quantitative discovery. We focus on the latter two stages introducing qualitative schemas as a means for representing qualitative discovery and showing how they can be learned by observation. IDS uses qualitative schemas to find quantitative laws. The implemented system has discovered intrinsic properties, such as the melting point of substances, and numeric laws, such as the conservation of mass of an object going through a phase change.

1. Introduction

Previous research in scientific discovery has shown that scientific discovery is a formidable task of machine learning. BACON (Langley, 1981) discovers a wide variety of numerical laws such as the ideal gas law. It also postulates intrinsic properties of object classes such as atomic weight. ABACUS (Falkenhainer, & Michalski, To appear) is similar to BACON but employs a better search mechanism to find numeric laws in a more efficient manner. It also improves upon BACON by identifying qualitative preconditions of quantitative laws. GLAUBER (Langley et.al., 1986), another scientific discovery system, addresses a different issue of empirical discovery, the formation of object taxonomies.

While each of these systems is successful at their task, they only address parts of the overall task in empirical discovery (Langley & Nordhausen, 1986). We are developing an integrated discovery system (IDS) which addresses many tasks of empirical discovery. IDS begins with the formation of object taxonomies, continues with qualitative discoveries, and then moves on to postulate numeric laws. There is not a strict ordering on these tasks. Qualitative discoveries usually lay the foundations for quantitative discoveries though the latter can lead to higher level qualitative discoveries and the formation of new object classes. In this paper, we focus on the stages of qualitative and quantitative discovery.

IDS operates in a simulated world of simple physics and chemistry. It thus overcomes one deficiency of previous discovery systems. Most of the existing discovery systems are provided with data and cannot perform experiments, an essential part of the discovery process.* IDS interacts with the world through a set of *effectors* and *sensors*. Using an effector, the system can actively alter an object's attribute e.g., by changing its location or heating it. Sensors let IDS inspect attributes of an object, such as temperature and mass. To carry out an experiment, the system applies an effector to some objects and observes the changes of the objects through a set of sensors.

The following section introduces the representation used in IDS for qualitative discoveries. The third and fourth sections detail the qualitative and quantitative discovery stages of IDS. We conclude with a description of the state of this research.

2. Representing Qualitative Schemas

Qualitative discovery involves the learning of the behavior of objects. Suppose we have some limited knowledge of what happens when we heat an object, e.g. we expect the temperature of the object to increase. If we actually heat a solid, we see that initially its temperature increases, but after sometime we may also observe the appearance of a new liquid object. At this point the temperature increase stops, and the mass of the liquid increases while the mass of the solid decreases. When the solid has disappeared, the temperature of the liquid begins to increase. This process continues until a gaseous object appears. As before the mass of the gas increases while the mass of the liquid decreases, the temperature of both objects remains constant during this period. Finally, the liquid vanishes and the temperature of the

---

* AM (Lenat, 1977) collects its own data and design its own experiments. But the mathematical domain of AM possesses features not easily extendable to "real world" domains.

gas increases, but so does its pressure. If we repeat this experiment and develop a model of this process with its distinct phases, we have learned how objects behave if they are heated. IDS is able to make the same kind of qualitative discovery generating *qualitative schemas*.

Our representation of qualitative behavior has been influenced by Forbus' *qualitative process* (QP) theory (1974). Qualitative schemas are similar to envisionments in QP theory. The schemas are finite state diagrams representing the behavior of objects over time. States correspond to distinct phases of an event during which objects exhibit constant behavior over time. Links identify successor states and the conditions which must be satisfied to go from the current state to a successor state. A state is represented as a frame with three slots: description, quantity-conditions, and process. The first two slots constitute the preconditions of a state. The description slot includes one or more classifications of the objects present in the state, for example solid or acid.* The slot also includes structural descriptions, such as heater h touches object a, or container a is connected to container b. The quantity conditions contain statements about attributes of the objects in the state. These statements are expressed as equalities or inequalities between the quantities of attributes and *limit-points* (see below). For example, the quantity conditions in a *melting* state of a heat-schema would identify that the temperature of the solid is equal to the melting point of the solid. The process is a list of zero or more *changes* which are occuring. We express a change in terms of the derivative of the changing attribute. For example, an increase in mass of object a is denoted $\Delta mass(a) > 0$.

A state ends only if the agent intervenes, for example by turning off the heat, or if the process reaches a limit-point such as the melting-point. Limit-points are importance because they are used in the quantity-conditions, and they form the basis for quantitative discoveries. The graphical illustration of a heat schema with the quantity-conditions and the process of each state is shown Figure 1. This example of the heat schema is strictly sequential; however, in general qualitative schemas include branches, i.e., there is more than one successor state for a given state.

The purpose of envisionments in QP theory and qualitative schemas in IDS differ. Qualitative physics is concerned with the explanation of physical behavior. In contrast, empirical discovery interprets observable behavior but does not provide an explanation for the behavior. IDS uses schemas to represent behavior of objects and to determine their properties. Furthermore, the knowledge of envisionments are encoded into systems using QP theory, while IDS learns qualitative schemas by observation. In the following section we describe this learning process.

3. Inducing Qualitative Schemas

IDS begins with a simple qualitative schema for each of its effectors. For example, the initial schema of the heat effector consists of two states: *s0*, with one object and no active process, and *s1*, with an object and a heater, where the heater touches the object. The temperature of the object in state *s1* is increasing. This represents IDS' initial knowledge of

---

* IDS forms classes of objects during the first stage of the discovery process. This stage is currently not implemented and a taxonomy is given to the system.
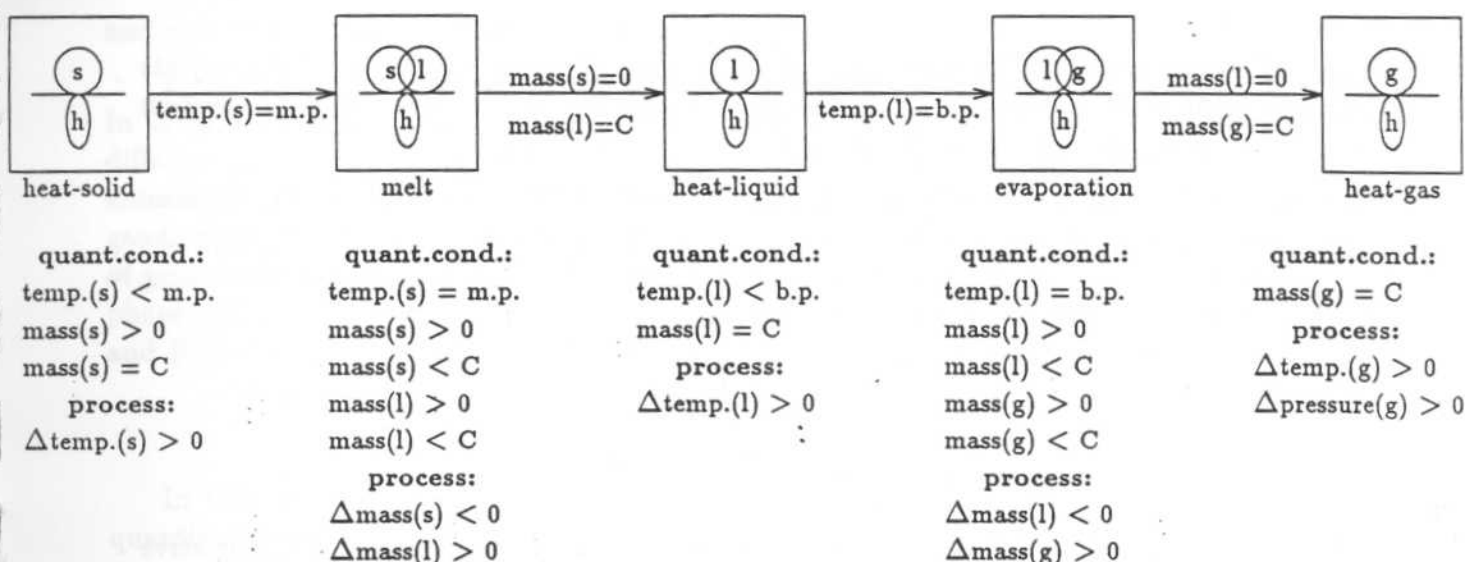
2

heat-solid	melt	heat-liquid	evaporation	heat-gas

The boxes are connected by arrows:
- heat-solid → melt: temp.(s)=m.p.
- melt → heat-liquid: mass(s)=0, mass(l)=C
- heat-liquid → evaporation: temp.(l)=b.p.
- evaporation → heat-gas: mass(l)=0, mass(g)=C

heat-solid

quant.cond.:
temp.(s) < m.p.
mass(s) > 0
mass(s) = C
process:
Δtemp.(s) > 0

melt

quant.cond.:
temp.(s) = m.p.
mass(s) > 0
mass(s) < C
mass(l) > 0
mass(l) < C
process:
Δmass(s) < 0
Δmass(l) > 0

heat-liquid

quant.cond.:
temp.(l) < b.p.
mass(l) = C
process:
Δtemp.(l) > 0

evaporation

quant.cond.:
temp.(l) = b.p.
mass(l) > 0
mass(l) < C
mass(g) > 0
mass(g) < C
process:
Δmass(l) < 0
Δmass(g) > 0

heat-gas

quant.cond.:
mass(g) = C
process:
Δtemp.(g) > 0
Δpressure(g) > 0

Figure 1: Qualitative Schema for heating an object*

the result of applying the heat effector to an object.

The system carries out experiments to improve its schemas, which can be refined in three ways. First, new states can be added. If IDS encounters behavior not seen before, it creates a new state and adds it to the schema. Second, two existing states not linkes before can be connected with new limit-points, indicating the conditions required to go from one state to the other. Finally, if IDS decides that the description of a state in a schema is too general, it augments the description of all states within that schema. A state description is too general if there is another state with the same preconditions but a different process.

Consider again the discovery of the heat schema. If IDS applies the heat effector to a block of ice, the temperature of the ice increases, satisfying all conditions of state *s1*. When the temperature of the ice reaches the melting point, a new object (liquid water) appears, and the mass of this new object increases, while the mass of the ice decreases. IDS' heat schema does not yet contain a state for this behavior. This causes IDS to create a new state *s2* and to add it to the schema. This state *s2* has a heater and two objects, *a* and *b*. The process slot describes the behavior of the objects, the decrease in mass of object *a* and the increase in mass of object *b*. After the ice disappears, state *s1* again accurately describes the current behavior, since the description does not yet distinguish between the types of objects. When the liquid reaches the boiling point, state *s2* now adequately describes the current behavior, so the schema is not changed at this point, i.e. IDS thinks it has found a loop. When the liquid disappears the preconditions for state *s1* are met again: there is a heater and an object, the gaseous water, and the heater touches the object. However, the process differs from the process of state *s1*. Not only does the temperature of the object increase, but so does its pressure. This violates one of IDS' basic assumptions, that under the same preconditions there cannot be two different processes. Thus the description of the state is

---

* Due to space limitations the start and end states are not shown.

too general. After some further experimentation, IDS decides that the phase of the objects (solid, liquid, or gas) should be included in the description of the states, and the system induces the heat schema as shown in Figure 1.

One can think of this schema-building process as searching the space of possible schemas. In these terms adding states and links make schemas more general, while augmenting the state description makes them more specific. The system also has a limited backtracking capability by merging states to avoid overly specific schemas.

4. Discovering Quantitative Laws

4.1 Experimentation

The discovery of numeric laws is the third stage of IDS' discovery sequence. Qualitative schemas serve two main purposes in this quantitative discovery stage. First, they provide a context for numeric laws. The ideal gas law in the context of a qualitative schema provides much more information than just the mathematical abstraction of the law which was all BACON found. A schema not only describes the applicability of a law but also specifies the pre- and post-conditions. Second, qualitative schemas constrain the search for numeric laws.

IDS runs experiments on objects to collect data about those objects. Most of the data for numeric laws is not directly observable but gathered in the form of limit-points and state duration. This information is recorded as attribute-value pairs. The classification tree which IDS forms in the first stage of its discovery process directs these experiments. For example, to learn more about a specific class such as water or acids, IDS applies several effectors invoking different schemas to all members of that class. To learn more about a particular schema, the system applies the effector associated with that schema to members of different classes. IDS uses the collected data to discover numeric laws in a manner similar to BACON and ABACUS.

4.2 Investigation

IDS takes a set of objects and their numeric attributes and defines numeric terms using mathematical functions such as addition and multiplication. Like BACON, the system discovers a quantitative law when it finds a constant numeric term. For example, in discovering the ideal gas law, IDS defines the term $X = PV/T$, where P is the pressure, V the volume and T the temperature of a gas. Since this term is constant for all gases, the system postulates a law that summarizes they regularity.

The quantitative laws that IDS discovers can be divided into two types. If the set of objects investigated is a class defined during the first phase of the discovery process (e.g. water or acids) a constant numeric term corresponds to an intrinsic property (Langley, 1981). Consider the class of water; all instances of this class have the same value for the boiling-point, this is the limit-point for the transition of the heat-liquid state to the boiling state. IDS thus asserts an intrinsic property associated with the class of water rather than with the specific instances of that class. More complex intrinsic properties include specific heat

and rate of reactions which combine several attributes.

The second type of quantitative discovery IDS finds are laws such as the ideal gas law. In order to postulate this type of law, the system searches for a constant term involving different objects appearing within the same instance of a schema. For example, the system notices that for all instances of the heat schema, the masses of the solid, the liquid and the gas appearing within the same instance of a schema are equal. Thus IDS postulates the law of conservation stating that the mass of an object remains constant as it goes through a phase change. Black's law of heat exchange, the conservation of mass in a chemical reaction, and Proust's law of constant proportions are also discovered in this manner.

5. Concluding Remarks

In this paper we have focused on those aspects of IDS that discover qualitative and quantitative laws using a heat-schema to illustrate the power of qualitative schemas. IDS also successfully builds a schema for combining two objects. including states for chemical reaction, and states observing Black's law of heat exchange, a schema for the fluid-flow of two connected containers filled with liquids (Forbus, 1984), and a schema for the osmosis of two liquids with different concentration (Rajamoney et.al., 1985). The system uses the schemas to run experiments on various objects in the simulated world, and the results of these experiments are used to discover numeric laws. IDS finds intrinsic properties such as the boiling and melting points of substances. It also discovers that the zero mass is a critical point for all objects since this is the point when objects appear and disappear. The numeric laws IDS discovers are conservation of mass in the heat process, equilibrium in fluid- and heat-flow and the osmosis. In its present form IDS can only discover qualitative laws which involve a simple constant term. Thus our next step is to incorporate a sophisticated search mechanism such as the ones used in BACON and ABACUS to discover numeric laws involving complex numeric terms. However, the large number of objects and attributes require an improved mechanism to handle irrelevant data.

Currently IDS is supplied with a taxonomy of objects but we are actively extending the system to construct taxonomies on its own initiative. Unlike previous clustering algorithms (Fisher & Langley, 1985), IDS will use symbolic attributes, numeric attributes, and information derived from qualitative schemas. It will start constructing classes based on observable attributes such as color and taste and uses these preliminary classes to build schemas. The behavior of objects within a schema and quantitative laws will then be used to reinforce existing classes or form new classes. For example, after forming a schema for reaction, IDS uses this information to form classes such as acids and bases.

As the capabilities of IDS grow, so will the need for an improved agenda mechanism which directs not only the three stages of the discovery process but also experimentation. This mechanism will have some of the characteristics of AM's agenda. The current prioritizing method is sufficient for the qualitative and quantitative discovery stages and enables IDS to produce interesting results.

6. References

Falkenhainer, B. C. and Michalski, R. S. (To appear). Integrating quantitative and qualitative discovery: The ABACUS system. *Machine Learning.*

Fisher, D. and Langley, P. (1985). Approaches to conceptual clustering. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* (pp. 691–697). Los Angeles, CA: Morgan Kaufmann.

Forbus, K. D. (1984). Qualitative process theory. In D. G. Bobrow (Ed.), *Qualitative Reasoning about Physical Systems,* (pp. 85–168). Cambridge, MA: MIT Press.

Langley, P. (1981) Data–Driven Discovery of Physical Laws. *Cognitive Science, 5,* 31–54.

Langley, P. and Nordhausen, B. (1986) A framework for empirical discovery. In *Proceedings of the International Meeting an Advances in Learning,* Les Arc, France.

Langley, P., Zytkow, J., Simon, H. A., and Bradshaw, G. L. (1986) The search for regularity: Four aspects of scientific discovery. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach, Vol. II* (pp. 425–469). Los Altos, CA: Morgan Kaufmann.

Lenat, D. B. (1977) Automated theory formation in mathematics. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence* (pp. 833–841). Cambridge, MA: Morgan Kaufmann.

Rajamoney, S., DeJong G., and Faltings, B. (1985). Towards a model of conceptual knowledge acquisition through directed experimentation. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* (pp. 688–690). Los Angeles, CA: Morgan Kaufmann.