Lecture 2 – Knowledge Representation

1. Knowledge Representation
   (a) What is it?
2. Logic and AI
   (a) As a Tool for Analysis and Specification
      i. An Example: Diagnosis
   (b) As a KR Language
      i. John McCarthy – Advice Taker.
      ii. Epistemological and Heuristic Adequacy
3. Davis' Framework

Knowledge Representation

1. Why represent knowledge?

2. What is knowledge?

3. What type(s) of representation should be used?

Why Represent Knowledge?

Aim of AI Research: Design and creation of systems capable of succeeding at tasks whose accomplishment requires intelligence.

Intelligent behavior is dependent on:

1. Massive amount of facts about environment, agent and capabilities.

2. Commonsense knowledge for incomplete situations, rules of thumb etc.

3. Different modes of reasoning: Deduction, Abduction, Induction. To derive new facts from old.

4. Biological constraints acquired through evolution and via agents symbiosis with the environment.

   Situated activity, Brooks, Maturana, Winograd, etc.

Require knowledge represention of:

1. Environment (Physics, specific domains etc.)

2. Agent and it's capabilities

3. Strategies for using knowledge in an efficient way. (Control)

In order to explain behavior of an intelligent agent given some goal, we attribute to the agent:

1. Knowledge of the conditions relevant to the success of the behavior.

2. Ability to reason from such knowledge.

Our Problem: Given a goal, how may an agent use facts about actions and their effects in a knowledge base to determine appropriate actions to acheive that goal?

1. How to represent actions and their effects in a knowledge base.

2. How to reason with such facts.

<u>Relation to Philosophy and Psychology:</u>

– Triviality of problem statement?
– Payoff in studying simple types of problems.
– Walking before running.

What is Knowledge?

– Philosophically an extremely subtle question. Has been discussed for thousands of years.
– **Justified True Belief**

In Artificial Intelligence:

– The things believed, known, or assumed are called *propositions*.
– Specific Properties:

1. Propositions are truth valuable. True, false, unknown, both. Truth values can be assigned to them.

2. Logical relations can hold between them, in particular logical consequence.

– A set $P_1, \ldots, P_n$ may be a good reason for believing Q.
– We can infer Q on the basis of $P_1, \ldots, P_n$.
– Neutrality as to choice of consequence.

Modeling an agent's beliefs:

– Assumption: at any one time, an agent has a set of beliefs(propositions).
– An *cognitive state*(its set of beliefs) is characterized in terms of it accepting as true *sentences* that express its propositions.

Knowledge is defined as propositions and propositions have syntactic representations in the form of well-formed formulas in a logical language.

Comments

– View as a loose engineering metaphor.

– Discuss the relation to philosophy and psychology. In particular new trends:

- Situated activity
- Imagery
- Neural nets
- Hard and soft AI
- Searle and the chinese room.

What Type(s) of KR should be used?

Our Choice: LOGIC!

Alternatives:

– Production Rules
– Semantic Nets
– Frame-based Languages
– Neural Nets

– Choice is highly controversial.

– Discuss Pros and cons.

Logic, AI and KR

Two General Views:

1. As a Tool for Analysis and Specification
   (a) As a source of analytical tools and techniques which provide the underlying mathematical and conceptual framework for AI research.

2. As a KR Language
   (a) As a source of languages and logics for artificial reasoners.

View 1: Logic as a Tool

The Diagnosis Task

– What is a diagnosis problem?

– How do we know that we have a solution to a diagnosis problem?

Goal: Logical definition of a diagnosis problem and a precise logical specification of a solution to such a problem.

We have a diagnosis problem when a device or system is observed to deviate from its desired specification.

Assumption 1 – There is a clear and precise specification of a system and its components.

Definition 1 *A* system *is a pair* $\langle SD, COMP\rangle$ *where* SD*, the* system specification*, is a satisfiable set of first-order sentences and* $COMP$ *is a set of constants, one for each distinguished component of the system.*

Definition 2 *An* observation OBS *is a finite set of first-order sentences.*

– Signature contains a non-logical unary predicate constant Ab whose intended meaning is 'abnormal'.

– $COMP$ is also a subset of the signature. We will use $\{c_1, \ldots, c_n\}$ as component constants.

Properly functioning system:

The following set of sentences is satisfiable

$$SD \cup \{\neg Ab(c_1), \ldots, \neg Ab(c_n)\}. \qquad (1)$$

Diagnosis Problem:

Let OBS be an observation of a set of faults. We have a diagnosis problem when the following set of sentences is unsatisfiable.

$$SD \cup \{\neg Ab(c_1), \ldots, \neg Ab(c_n)\} \cup OBS. \qquad (2)$$

– A diagnosis is going to be a *conjecture* (we can not know for sure) that certain of our components are faulty (Ab) and the rest are normal ($\neg Ab$).

– What components do we retract from $\{\neg Ab(c_1), \ldots, \neg Ab(c_n)\}$ to restore satisfiability?

Here is Reiters formal definition.

Definition 3 *A* diagnosis *for* $\langle\langle SD, COMP\rangle, OBS\rangle$ *is a minimal set* $\Delta \subseteq COMP$ *such that*

$$SD \cup OBS \cup \{Ab(c) \mid c \in \Delta\} \\ \cup\{\neg Ab(c) \mid c \in COMP - \Delta\} \qquad (3)$$

is consistent.

Based on the minimality of Δ , it follows that for each $c_i \in \Delta$,

$SD \cup OBS \cup \{\neg Ab(c) \mid c \in COMP - \Delta\} \models Ab(c_i)$.

This states that the faulty components are logically dependent on the functioning components in $COMP - \Delta$.

Note the value of the logical specification above:

– Reiter's logical analysis of diagnostic problems and their solutions could be used as a description of a minimium criterion of success for any implemented diagnosis system.

– A diagnosis system would be deemed successful if for a particular diagnosis, it returned at least one of the minimal sets of components that the logical representation of the same problem would.

– One could implement the diagnosis system by constructing a theorem prover that generates the minimal sets of faulty components

– The logical analysis is completely independent of questions regarding implementation.

Binary Full-adder Example

– Diagnosis of a binary full adder.
– Components: $COMP = \{A_1, A_2, X_1, X_2, O_1\}$.

System specification SD:

$$\begin{aligned}\forall x. AndG(x) \wedge \neg Ab(x) \rightarrow out(x) = \\ and(in1(x), in2(x)).\end{aligned} \tag{4}$$

$$\begin{aligned}\forall x. XorG(x) \wedge \neg Ab(x) \rightarrow out(x) = \\ xor(in1(x), in2(x)).\end{aligned} \tag{5}$$

$$\begin{aligned}\forall x. OrG(x) \wedge \neg Ab(x) \rightarrow out(x) = \\ or(in1(x), in2(x)).\end{aligned} \tag{6}$$

$$\begin{aligned}\forall x. OrG(x) \wedge \neg Ab(x) \rightarrow out(x) = \\ or(in1(x), in2(x)).\end{aligned} \tag{7}$$

$$AndG(A_1) \wedge AndG(A_2). \tag{8}$$

$$XorG(X_1) \wedge XorG(X_2) \wedge OrG(O_1). \tag{9}$$

$$out(X_1) = in2(A_2). \tag{10}$$

$$out(X_1) = in1(X_2). \tag{11}$$

$$out(A_2) = in1(O_1). \tag{12}$$

$$in1(A_2) = in2(X_2). \tag{13}$$

$$in1(X_1) = in1(A_1). \tag{14}$$

$$in2(X_1) = in2(A_1). \tag{15}$$

$$out(A_1) = in2(O_1). \tag{16}$$

Circuit Inputs are Binary Valued:

$$in1(X_1) = 0 \lor in1(X_1) = 1. \tag{17}$$

$$in2(X_1) = 0 \lor in2(X_1) = 1. \tag{18}$$

$$in1(A_1) = 0 \lor in1(A_1) = 1. \tag{19}$$

In addition there are some axioms for boolean algebra not specified here.

Full Binary Adder Diagram

Inputs and Outputs (Observations):

– Suppose the adder is given the inputs 1, 0, and 1 and it outputs 1, 0 in response.

– This can be described as the following observation OBS:

$$in1(X_1) = 1 \wedge in2(X_1) = 0 \wedge in1(A_2) = 1. \qquad (20)$$

$$out(X_2) = 1 \wedge out(O_1) = 0. \qquad (21)$$

– Note that the observations tell us that the adder is faulty.

– Both circuit outputs are wrong for the given inputs.

Diagnosis:

– Three minimal sets of components:

$$\{X_1\} \text{ or } \{X_2, O_1\} \text{ or } \{X_2, A_2\}$$

Logic as a KR Language

– McCarthy[1959] “Programs with Common Sense”. The Advice Taker.

> The “Advice Taker” is a proposed program for solving problems by manipulating sentences in formal languages.
>
> The main advantages we expect the advice taker to have is that its behavior will be improvable merely by making statements to it, telling it about its symbolic environment and what is wanted from it. To make these statements will require little if any knowledge of the program or the previous knowledge of the advice taker. One will be able to assume that the advice taker will have available to it a fairly wide class of immediate logical consequences of anything it is told and its previous knowledge. This property is expected to have much in common with what makes us describe humans as having common sense. We shall therefore say that a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows.

– a research strategy aimed at the design, implementation and analysis of artificial reasoners.
– Discuss extraordinary vision and the ambitious program.

> "ultimate objective is to make programs that learn from their experience as effectively as humans do", but "in order for a program to be capable of learning somthing it must first be capable of being told it".

Emphasis on the following feature of a system that will evolve intelligence of human order:

1. Interesting changes in behavior must be expressible in a simple way.

The criterion for success will be to

> "achieve a system which can be told to make a specific improvement in its behavior with no more knowledge of its internal structure or previous knowledge than is required in order to instruct a human".

–Viewed from this context, logical languages seem to be the ideal choice as a knowledge representation language. Why is this?

McCarthy lists the advantages of the use of declarative sentences:

1. Advantage can be taken of previous knowledge.

2. Declarative sentences have logical consequences and it can be arranged that the machine will have available sufficiently simple logical consequences of what it is told and what it previously knew.

3. The meaning of declaratives is less dependent on the order in which they are presented than is the case with imperatives.

4. The effect of a declarative is less dependent on the previous state of the system so that less knowledge of this state is required on the part of the instructor.

The KR Hypothesis

– Modern Variant due to B. Smith.

> Any mechanically embodied intelligent process will comprise structural ingredients that a) we as external observers naturally take to represent a propositional account of the knowledge that the overall process exhibits, and b) independent of such external semantical attribution, play a formal but causal and essential role in engendering the behavior that manifests that knowledge.

Objections to the Logic-Based Approach

Israel lists three:

1. The alleged, or real, expressive inadequacies of logical languages.

2. The difficulty of theorem-proving, or finding derivations.

3. The nonmonotonicity or defeasibility of much of ordinary reasoning.

Replies:

1. Although unnatural to humans, not necessarily unnatural to machines.
   Logic is not FOPC!

2. Tradeoff between expressive power and computational tractability. Undecidability and efficiency. Find fragments. Specialized languages.

3. Noticed by proponents themselves. Extensions of logic proposed. Makes (2) even more critical.

Epistemological vs. Heuristic Adequacy

- Epistemological Part of AI studies
  - What kinds of facts about the world are available to an observer with given opportunities to observe,
  - how these facts can be represented in the memory of a computer,
  - and what rules premit legitimate conclusions to be drawn from these facts.
- Heuristic Part of AI studies issues such as
  - how to search spaces of possibilities
  - how to pattern match, etc.

Advantages of the Separation

1. Same problem of what info is available and what conclusions can be drawn from it are similar for many reasoning tasks.

2. Single solution to epistemological problem can support a variety of heuristic approaches to the problem.

3. AI is a difficult scientific problem. Divide and conquer is a good methodology.

4. Difficult to formalize facts of common knowledge. Confining oneself only to special cases doesn't face difficulties involved in general intelligence.

Reichgalt's 4 levels of KB-systems

4 levels and criteria for evaluating:

Implementation – Efficiency.

Logical – Clear syntax and semantics with sound inference rules.

Epistemological – Naturalness with which a representation can be constructed and understood and its modularity. Small additions imply small changes to the KB.

Conceptual – Direct correlation with phenomena modeled.

A KR Framework (Davis)

– *Commonsense Reasoning* is the performance of *inference* on a body of *object-level* information, which describes the subject matter of the reasoning.
–*Control-level* information describes the internal state of the reasoner and is used to control the process of inference.

Knowledge-Based Systems will be defined as a program consisting of two parts:

- *Knowledge Base* – A data structure encoding a body of object-level information.

- *Knowledge base manager* – Collection of procedures that perform inferences on information in the KB, including procedures that add and delete information.

– As much information as possible should be encoded declaratively.

– Possible modes of inference:

1. **Deduction** – Logically sound inferences. Inferred info. necessarily true if premises are true.

2. **Abduction** – Offer plausible explanations for some particular fact. Non-deductive. Useful for explanation, planning, diagnosis.

3. **Induction** – Infers general rules from specific examples. Non-deductive. Useful for learning.

Predominant Methodology

– in Commonsense-Reasoning and Knowledge Representation:

> . . . is to take a number of examples of commonsense inference in a commonsense domain, generally deductive inference; identify the general domain knowledge and the particular problem specification used in the inference; develop a formal language in which this knowledge can be expressed; and define primitives of the language as carefully and precisely as possible.

– **Method of Representative Examples**

Davis lists three methodological assumptions:

1. In designing a reasoning system, it is generally worthwhile to characterize the object level information that the system will use and the inferences that it will perform, independently of determining how the system should perform these inferences.

2. A number of fundamental commonsense domains arise so frequently in designing reasoning systems that it is worthwhile for the research community to study the basic structure of commonsense inference in these domains independently of any particular applications.

3. Certain characteristics of knowledge used in commonsense inference make it particularly important that the language used to express this knowledge be very carefully and tightly defined. In particular:

   (a) Commonsense inferences tend to use a great variety of partial and incomplete knowledge. Representations left open for interpretation may be ambiguous or unclear to the programmer.

   (b) The very familiarity of commonsense concepts and their closeness to natural language forms renders them liable to ambiguity.

Designing a Commonsense Reasoning System

– "a commonsense reasoning system be designed as a knowledge base, recording information and performing inference on object-level information about some *microworld*".
– A microworld is defined as

"a restricted, idealized model of the world containing only those relations and entities of interest in the particular reasoning system being designed."

In addition, the microworld is analyzed on three levels:

1. The definition of the microworld: its entities, its relations, and the rules that govern them. This definition is called the *ontology* of the domain.

2. The analysis of the type of information that the system must record and the types of inferences it must perform. The result of this analysis is generally a *formal language*, in which facts of the microworld can be expressed, and *axioms* or *rules of inference*, which characterize inferences that are reasonable within the microworld.

3. An *implementation* of the information in mutable data structures and of the inferences in procedures and predefined data structures.

– Ontology, Syntax and Semantics, Implementation

Diagram of Davis Framework