Explanation-Based Learning (EBL)

[Mitchell et al. 86; DeJong and Mooney 86]

Learning:

- Directed towards a goal
- Guided by a domain theory

Why EBL is needed

- What's a good generalization of:

  "On July 18, Mike put on his Hawaiian shirt and went to Paris by driving from Bloomington to Indianapolis and flying to Paris."

Some inductive generalizations of "On July 18, Mike put on his Hawaiian shirt and went to Paris by driving from Bloomington to Indianapolis and flying to Paris:"

- On July 18, people wearing Hawaiian shirts in Bloomington go to Paris by ...
- People in Bloomington go to Paris by ...
- In summer, people in Bloomington go to Paris by ...
- People vacation in the summer.
- All trips start from Bloomington.
- People everywhere going anywhere go to Indianapolis first.
- To go short distances, people drive.
- In hot weather, people wear Hawaiian shirts.

- Inductive learning requires many examples
- In real-world situations, situation identification is a serious problem (which examples are relevant—travel stories? summer stories? fashion stories? stories about Mike? ...)

EBL learns from single examples, focusing using

- Domain theory
- Training example
- User-supplied goal (called the "target concept")

One motivation: Learning by observation

Example:

- Domain theory: General world knowledge.
- Training example: "On July 18, Mike put on his Hawaiian shirt and went to Paris by driving to Indianapolis and flying to Paris."
- User-supplied goal to select features: Explain Mike's clothing selection.

The explanation involves: In hot weather, Hawaiian shirts are a good choice . . .

Requiring a target concept: Is it cheating?

Yes and no.

No: Doesn't preclude valuable (and hard) learning. Target concepts could be:

- Why Microsoft makes so much money
- Why an airplane flies
- Why a song prompts an emotion

Yes: Autonomous systems need to decide for themselves.

The EBL algorithm

- Input target concept and training example
- Build deductive proof of why the example belongs to concept
- Generalize from the proof and learn generalization

The classic EBL example:
Learning to recognize mugs

- Domain theory: rules about containment, insulation, finger motion, etc.
- Training example: your mug
- User-supplied goal:
  Using mug as container for hot liquids

Training example:

Target concept: Good mugs satisfy
Graspable(X) and Insulates(X) and Holds-Liquid(X)

Proof could use rules such as:

Has-handle(X) $\Rightarrow$ Graspable(X)

Ceramic-material(X) $\Rightarrow$ Insulates(X)

Concave-up(X) $\Rightarrow$ Holds-Liquid(X)

Conclusion: Cups are ceramic objects with upward concavity and handles

What's being learned?

- EBL systems already have a correct concept definition: The target concept.

  E.g., cups are graspable, insulate and hold liquid.

- New definition is more restrictive.

What's the point?

Operationality: EBL puts the definition into more usable form.

- Features in first cup definition can't be directly observed. Features in the second can.

- Features in "flying machine" don't help to build one. Component descriptions do.

What's the role of the target example?

- EBL assumes a perfect domain theory; example doesn't make generalization more correct.

- But example may help to show which specializations are important in the world.

Summary

- EBL uses a theory and target concept to do single-example learning.

- EBL is *knowledge transformation*.

What are the problems?

- Explanation is expensive

- What if the theory is incomplete or wrong?

  A robot hand won't match its idealization . . .

- Where do target concepts come from?