fletcher-scott.com

MACHINE TEACHING CAUSAL PROPOSITIONS ACROSS MODEL EQUIVALENCE CLASSES

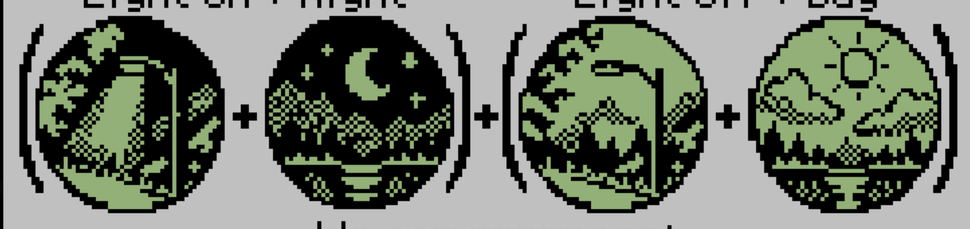
FLETCHER SCOTT
DANULA HETTIACHCHI
LAUREN SALING
DAMIANO SPINA

The problem

To decide whether a statement is true or false, a learner has to work out how what they observe can be used to predict what would happen in their environment if something were deliberately changed.[1]

For example, I tell my robot, "Streetlights make it dark."
It cannot turn the streetlights on or off or change the time of day. It can only observe that streetlights are on when it is dark and use those observations to decide what would happen if the streetlights were deliberately turned on during the day.[1,2]

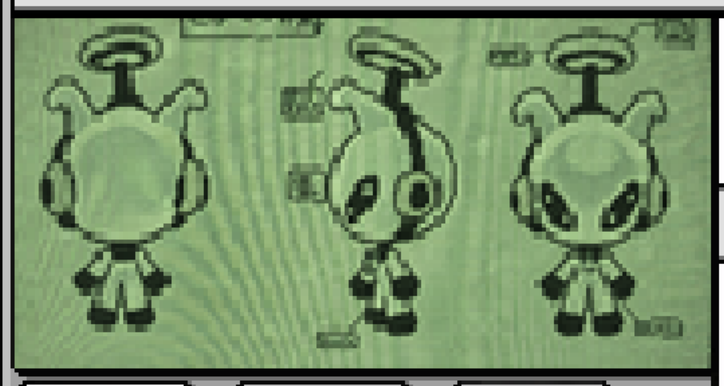
Light On + Night Light Off + Day



We can represent observations as basic symbols!

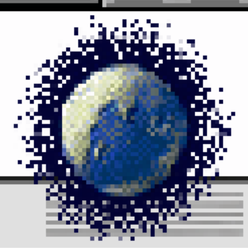
R M I T S H B S

Does this remind you of anything?



"Carbon dioxide does not cause global warming."
"Vaccines cause autism."
"Postal voting causes widespread voter fraud."

Click Here!



Task: Teaching the robot

Can we choose examples that make the robot more confident in the correct answer?

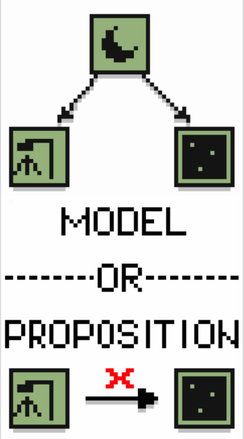
00% CONFIDENCE TRUE FALSE

options | model | prop.

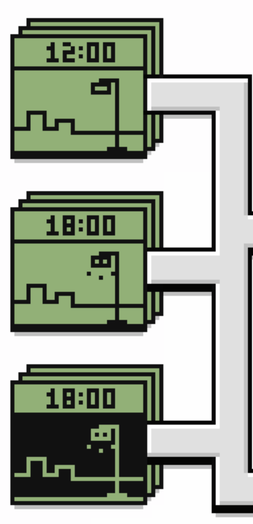
Two teaching strategies

- Teach the model
  - Show the robot examples that help it learn how the whole system works
- Teach the proposition
  - Show the robot examples that help it decide whether this particular statement is true/false.

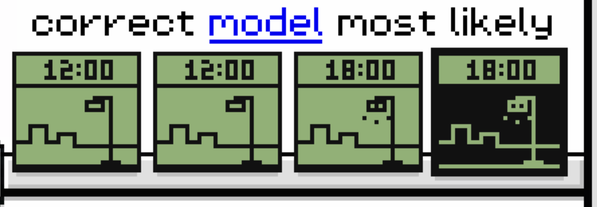
The World | The Observations | CHOOSING EXAMPLES | The Proposition



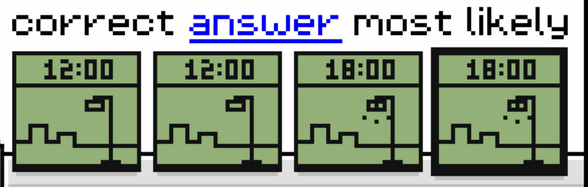
All possible observations from this world...

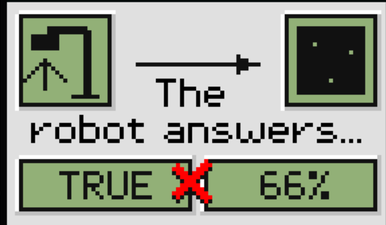


Pick four to make the correct model most likely



Pick four to make the correct answer most likely





PAPER

ADMSpedia

article | discussion | edit this page | history

The study

Machine teaching tries to find the best examples and the best order in which to present those examples, for helping a learner reach a goal.[3][4][5] This line of research has developed increasingly good ways to adapt teaching to the learner, either by learning more about them as teaching unfolds,[6][7][8] or by making the best possible use of what is already known about how they learn.[9][10][11] This has shown that what makes an example "best" depends partly on **how different learners use the same information**.[9][10][11] These approaches have successfully been used to improve teaching for human learners.[12][13] **But the learner is only one source of variation. The learning problem can change too.**

Comparing the teaching strategies

If we take a set of questions covering every possible direct cause, so that knowing all the answers reveals the model's structure, are the best examples for learning the model also best for answering those questions?
To test this, we compared teachers targeting each goal against random selection, using different computational learners with the same pool of observations.

Glossary

Structural causal model (SCM):
A model of what causes what, and how.

Equivalence class:
A set of different models that are cannot be told apart.

Command Prompt - Teaching results

```
C:\RMIT_Robot_Model_Inspection_Tool> compare_teaching.exe        19,935 SCMs / 4 learners / 24 questions
A shared endpoint did not mean the same examples were best along the way.

Changing the objective:
Proposition teaching generally favoured answers; model teaching favoured graph recovery. Gains did not necessarily transfer.

Changing the learner:
Proposition teaching led for 3/4 learners at 16 records; the simplicity-biased learner favoured model teaching. Order affected bounded learners, not cumulative learners after all records.

Changing the question:
Some questions resolved before the graph. Selected records helped even when the full distribution left the answer unresolved.

[01] SELECTION / 16 records / 24 questions
Lower mean error:  ■ Prop.  ■ Model  ▣ Tie
                                  DIRECT   PATH    MED.    CP      CC
Ideal Inference (II)              ■■■■■■   ■■■■■■  ■■■■■■  ■■■■   ■■■■
Prefers simpler explanations (SE) ■■■■■■   ■■■■■■  ■■■■■■  ■■■■   ■■■■
Remembers last four (R4)          ■■■■■■   ■■■■■■  ■■■■■■  ■■■■   ■■■■
Revises one graph at a time (RG)  ▣■▣■■■   ■■■■■■  ▣■■▣▣▣  ▣▣■■   ▣▣■■

[02] ORDER / same 32 records                                  Brier error
Lower is better  — Prop.  — Model  — Random
                 II                SE                R4                RG
0.5          ·                 ·                 ·                 ·
0.3          ·                 ·                 ·                 ·
0.1          ·                 ·                 ·                 ·
             4 8  16  32       4 8  16  32       4 8  16  32       4 8  16  32
                                                                    Records seen: 4 / 8 / 16 / 32

Best examples depend on the goal, learner and question.
```