Talk

Overview

- Introduction to Program Synthesis
    - Deductive Synthesis
    - Inductive Synthesis
- Formal Inductive Synthesis
- Probabilistic Inductive Synthesis
- Interesting problems

Formal Inductive Synthesis

- $\mathbf{E}$: The (finite/infinite) domain of examples.
  - A concept is a set of examples within $\mathbf{E}$. ie: $c \subseteq \mathbf{E}$.
  - An example is a specific instantiations from $\mathbf{E}$.
- $\mathscr{C}$: The concept class. A set of all possible concepts from $\mathbf{E}$. Also, $c \subseteq 2^{\mathbf{E}}$.
- Φ : The specification. A set of "correct" concepts.
  - $x \vdash \Phi$. A positive example x. ie: $x \in \mathbf{E}$ and $x \in c$ such that $c \in \Phi$.
  - In inductive synthesis, $\Phi \subseteq \mathbf{E}$ and c satisfies Φ iff $c \in \Phi$.
- **Formal Inductive Synthesis**: Given $\mathscr{C}$ and $\mathbf{E}$, find -- using a subset of $\mathbf{E}$ -- a target concept $c \in \mathscr{C}$ that satisfies specification $\Phi \subseteq \mathscr{C}$.
- $\mathscr{O}$: *Oracle Interface*. Subset of $\mathscr{Q} \times \mathscr{R}$ that are semantically well formed.
  - $\mathscr{Q}$: Set of query types.
  - $\mathscr{R}$: Set of response types.

Formal Inductive Synthesis

Generalization of Inductive Program Synthesis.

Definition: *Given a concept class* $\mathscr{C}$*, and a domain of examples* $\mathbf{E}$*, find - using a subset of* $\mathbf{E}$ *- a target concept* $c \in \mathscr{C}$ *that satisfies specification* $\Phi \subseteq \mathscr{C}$*.*

- $\mathbf{E}$: The (finite/infinite) domain of examples. $x \in \mathbf{E}$ is an input-output behavior.
- c : A concept is a set of examples within $\mathbf{E}$. $\mathscr{C}$ is the set of all possible concepts. $\mathscr{C} \subseteq 2^{\mathbf{E}}$.
- Φ : The set of correct concepts. $x \vdash \Phi$ if $x \in \mathbf{E}$ and $x \in c$ such that $c \in \Phi$.
    - In inductive synthesis, $\Phi \subseteq \mathbf{E}$ and c satisfies Φ iff $c \in \Phi$.

Oracle Guided Inductive Synthesis

- $\mathbf{O}$: Oracle. A non-deterministic mapping $\mathbf{O} : \mathbf{D}^* \times \mathbf{Q} \rightarrow \mathbf{R}$
  - $\mathbf{Q}$: Set of queries to the oracle.
  - $\mathbf{R}$: Set of responses from the oracle.
  - d: A valid dialogue pair $(q, r) \in \mathbf{Q} \times \mathbf{R}$ such that (q, r) *conforms* to an oracle interface $\mathscr{O}$.
  - $\mathbf{D}$: Set of valid dialogue pairs for an oracle interface. $\mathbf{D}^*$ denotes the set of valid dialogue sequences (of finite length).
- $\mathbf{L}$: Learner. A non-deterministic mapping $\mathbf{L} : \mathbf{D}^* \rightarrow \mathbf{Q} \times \mathscr{C}$
- An OGIS procedure $\langle \mathbf{O}, \mathbf{L} \rangle$ is said to solve an FIS $\langle \mathscr{C}, \mathbf{E}, \Phi, \mathscr{O} \rangle$ with dialogue sequence δ if $\exists i . \mathbf{L}(\delta[i]) = (q, c), c \in \mathscr{C} \wedge c \in \Phi$ and $\forall j > i \exists q' \mathbf{L}(\delta[j]) = (q', c)$ (ie: OGIS converges to c).

Complexity of OGIS

The complexity of OGIS depends on:

1. Learner complexity: The complexity of each invocation of $\mathbf{L}$
2. Oracle complexity: The complexity of each invocation of $\mathbf{O}$
3. Query complexity: The number of iterations of the OGIS loop before convergence.

Counter-Example Guided Inductive Synthesis

-

- (q_{ce}, r_{ce}): The queries ask for counterexamples to a particular concept ($c \notin \Phi$)

Counter-examples can contain vary-ing degree of information:

1. $\mathtt{CEGIS}$: Arbitary Conterexamples
2. $\mathtt{MINCEGIS}$: Minimal Counterexamples
3. $\mathtt{CBCEGIS}$: Constant-bounded counterexamples
4. $\mathtt{PBCEGIS}$: Positive-bounded counterexamples

The learner can have varying degrees of memory:

1. Infinite Memory
2. Finite Memory

Under Finite Memory

- 4 possible synthesis engines:
  - $T_{\mathtt{CEGIS}}$
  - $T_{\mathtt{MINCEGIS}}$
  - $T_{\mathtt{CBCEGIS}}$
  - $T_{\mathtt{PBCEGIS}}$
- Using arbitary counter-examples and minimal counterexamples

Type	Query	Response
$q_{\text{mem}}(x)$	Is x positive or negative?	Yes/No
q_{wit}^{+}	Give me a positive example.	$x \vdash \Phi$
q_{wit}^{-}	Give me a negative example.	$x \not\vdash \Phi$
$q_{\text{ce}}(c)$	Prove that c doesn't satisfy Φ	Proof / $\perp$
$q_{\text{corr}}(c)$	Prove that c satisfies Φ	Yes / No w/Proof

$$\mathscr{C} = \{\{1,2\},\{3,4\}\}$$
$$\Phi = \{\{1,2\}\}$$
$$\mathbf{E} = \mathbb{N} \times \mathbb{N}$$

Goal: Find a set of queries Q of minimum cardinality that recovers all concepts in $\mathscr{C}$.

- $TD(\mathscr{C})$: The minimum number of examples that a teacher must reveal to uniquely identify $c \in \mathscr{C}$.
- $M_{\text{OGIS}}(\mathscr{C}) \geq TD(\mathscr{C})$
  - If $M_{\text{OGIS}}(\mathscr{C}) < TD(\mathscr{C})$, there exists a shorter teaching sequence than $TD(\mathscr{C})$. Contradiction.
- $\frac{VC(\mathscr{C})}{\log(|\mathscr{C}|)} \leq TD(\mathscr{C}) \leq |\mathscr{C}| - 1$
  - Known result from literature on teaching dimension.