Programmatic Strategy Synthesis

Resolving Nondeterminism in Probabilistic Programs

KEVIN BATZ, RWTH Aachen University, Germany
TOM JANNIK BISKUP, RWTH Aachen University, Germany
JOOST-PIETER KATOEN, RWTH Aachen University, Germany
TOBIAS WINKLER, RWTH Aachen University, Germany

Markov decision processes (MDPs) are a standard model for sequential decision making and planning in probabilistic environments. A fundamental MDP problem is to find a decision rule — a *strategy* — that performs well according to a given criterion. In this talk, we are concerned with synthesizing strategies for infinite parameterized families of MDPs. For example, in a gambling game we may not know our initial budget upfront, and would like to have a *parametric strategy*, i.e, one for each possible initial budget. We encode such families of MDPs as well as the corresponding strategies as imperative programs extended with coin flips and nondeterminstic branchings. The latter are used to model the agent's decisions as can be seen in the example further below.

Our approach draws on principles from deductive program verification, most prominently pre- and postconditions as well as loop invariants. The technique yields strategies with mathematically provable performance guarantees and can be automated if all loops occurring in the program are annotated with suitable (user-provided) invariants.

This presentation is based on a paper published at POPL 2024: https://dl.acm.org/doi/10.1145/3632935.

$$
\begin{array}{l}
\texttt{while}\, c < N \land a = 0 \to \{\\
\quad \texttt{if true}\\
\qquad \to \{\, c := c + 1 \,\}\ [p]\ \{\, a := 1 \,\}\\
\quad \texttt{if true}\\
\qquad \to \{\, c := c + 2 \,\}\ [q]\ \{\, a := 1 \,\}\\
\}
\end{array}
$$

Program modeling a gamble which is won as soon as $c \geq N$ tokens are collected and lost once a is set to 1. The two "true" branches model a choice of the gambler. Statements of the form $\{\dots\}\ [p]\ \{\dots\}$ represent a probabilistic choice. The initial values of c, N, p, and q are arbitrary parameters.

$$
\begin{array}{l}
\texttt{while}\, c < N \land a = 0 \to \{\\
\quad \texttt{if}\ (q \leq p^2)\ \lor\ (p^2 < q < p \land N - c \textit{ is odd})\\
\qquad \to \{\, c := c + 1 \,\}\ [p]\ \{\, a := 1 \,\}\\
\quad \texttt{if}\ (p \leq q)\ \lor\ (q = p^2)\ \lor\ (p^2 < q < p \land N - c \geq 2)\\
\qquad \to \{\, c := c + 2 \,\}\ [q]\ \{\, a := 1 \,\}\\
\}
\end{array}
$$

Parametric strategy represented as a refined version of the program on the left. The gambler is guaranteed to win with highest possible probability if decisions consistent with the red predicates are made. This works for arbitrary initial values of c, N, p, and q.

Authors' addresses: Kevin Batz, RWTH Aachen University, Germany, kevin.batz@cs.rwth-aachen.de; Tom Jannik Biskup, RWTH Aachen University, Germany, tom.biskup@rwth-aachen.de; Joost-Pieter Katoen, RWTH Aachen University, Germany, katoen@cs.rwth-aachen.de; Tobias Winkler, RWTH Aachen University, Germany, tobias.winkler@cs.rwth-aachen.de.