Probabilistic Systems and Bisimulation

Prakash Panangaden[1]

[1]School of Computer Science
McGill University

Estonia Winter School March 2015

Outline

Summary of Results

- Probabilistic bisimulation can be defined for continuous state-space systems. [LICS97]
- Logical characterization. [LICS98,Info and Comp 2002]
- Metric analogue of bisimulation. [CONCUR99, TCS2004]
- Approximation of LMPs. [LICS00,Info and Comp 2003, QEST 2006, CONCUR 2004]
- Weak bisimulation. [LICS02, CONCUR02]
- Real time. [QEST 2004, JLAP 2003, LMCS 2006]
- Event bisimulation. [Info and Comp 2006]
- Applications to machine learning [UAI 2004-06, AAAI 2015]
- Approximation by Averaging [JACM 2014]

Labelled Transition System

- A set of states S,
- a set of *labels* or *actions*, L or $\mathcal{A}$ and
- a transition relation $\subseteq S \times \mathcal{A} \times S$, usually written

$$\rightarrow_a \subseteq S \times S.$$

The transitions could be indeterminate (nondeterministic).

Markov Chains

- A *discrete-time* Markov chain is a finite set S (the state space) together with a transition probability function $T : S \times S \longrightarrow [0, 1]$.
- A Markov chain is just a probabilistic automaton; if we add labels we get a PTS.
- The key property is that the transition probability from s to s' only depends on s and s' and not on the past history of how it got there. This is what allows the probabilistic data to be given as a single matrix T.

Discrete probabilistic transition systems

- Just like a labelled transition system with probabilities associated with the transitions.
- $$(S, \mathsf{L}, \forall a \in \mathsf{L}\ T_a : S \times S \longrightarrow [0, 1])$$
- The model is *reactive*: All probabilistic data is *internal* - no probabilities associated with environment behaviour.

Examples of PTSs

$$
\begin{tikzcd}
 & s_0 \arrow[dl, "a[\frac{1}{4}]"'] \arrow[dr, "a[\frac{3}{4}]"] & \\
s_1 & & s_2 \arrow[d, "a[1]"] \\
 & & s_3
\end{tikzcd}
$$

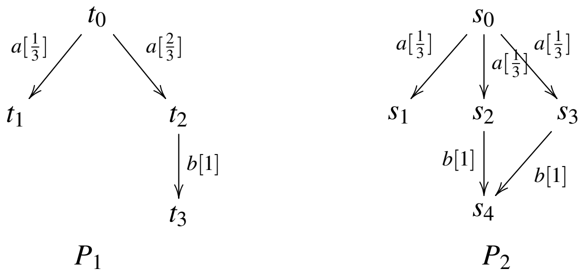
A_1

$$
\begin{tikzcd}
 & s_0 \arrow[dl, "a[1]"'] \arrow[dr, "b[1]"] & \\
s_1 \arrow[d, "c[\frac{1}{2}]"'] \arrow[drr, "c[\frac{1}{2}]"] & & s_2 \arrow[d, "a[1]"] \\
s_4 & & s_3
\end{tikzcd}
$$

A_2

Bisimulation for PTS: Larsen and Skou

- Consider



- Should s_0 and t_0 be bisimilar?
- Yes, but we need to add the probabilities.

The Official Definition

- Let $\mathcal{S} = (S, \mathsf{L}, T_a)$ be a PTS. An equivalence relation R on S is a **bisimulation** if whenever sRs', with $s, s' \in S$, we have that for all $a \in \mathcal{A}$ and every R-equivalence class, A, $T_a(s, A) = T_a(s', A)$.
- The notation $T_a(s, A)$ means "the probability of starting from s and jumping to a state in the set A."
- Two states are bisimilar if there is some bisimulation relation R relating them.

What are labelled Markov processes?

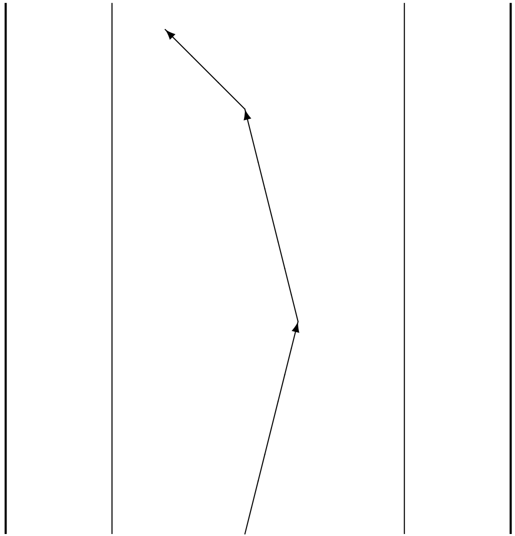
- Labelled Markov processes are probabilistic versions of labelled transition systems. Labelled transition systems where the final state is governed by a probability distribution - no other indeterminacy.
- All probabilistic data is *internal* - no probabilities associated with environment behaviour.
- We observe the interactions - not the internal states.
- **In general, the state space of a labelled Markov process may be a *continuum*.**

Motivation

Model and reason about systems with *continuous* state spaces or continuous time evolution or both.

- hybrid control systems; e.g. flight management systems.
- telecommunication systems with spatial variation; e.g. cell phones
- performance modelling,
- continuous time systems,
- probabilistic process algebra with recursion.

An Example of a Continuous-State System



a - turn left

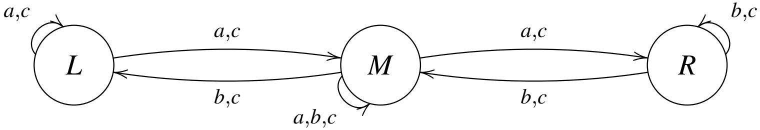
b - turn right

c - straight

Actions

a - turn left, b - turn right, c - keep on course
The actions move the craft sideways with some probability distributions on how far it moves. The craft may "drift" even with c. The action a (b) must be disabled when the craft is too near the left (right) boundary.

Schematic of Example

- 

- This picture is misleading: unless very special conditions hold the process cannot be compressed into an *equivalent* (?) finite-state model. In general, the transition probabilities should depend on the position.

Some remarks on the use of this model

- This is a toy model but exemplifies the issues.
- Can be used for reasoning - much better if we could have a finite-state version.
- Why not discretize right away and never worry about the continuous case? Because we lose the ability to *refine* the model later.
- A better model would be to base it on rewards and think about finiding optimal policies as in AI literature.

Recap of Markov Kernels

- A *Markov kernel* is a function $h : S \times \Sigma \longrightarrow [0, 1]$ with (a) $h(s, \cdot) : \Sigma \longrightarrow [0, 1]$ a (sub)probability measure and (b) $h(\cdot, A) : X \longrightarrow [0, 1]$ a measurable function.
- Though apparantly asymmetric, these are the stochastic analogues of binary relations
- and the uncountable generalization of a matrix.

Formal Definition of LMPs

- An LMP is a tuple $(S, \Sigma, \mathsf{L}, \forall\alpha \in \mathsf{L}.\tau_\alpha)$ where $\tau_\alpha : S \times \Sigma \longrightarrow [0, 1]$ is a *transition probability* function such that
- $\forall s : S.\lambda A : \Sigma.\tau_\alpha(s, A)$ is a subprobability measure
  and
  $\forall A : \Sigma.\lambda s : S.\tau_\alpha(s, A)$ is a measurable function.

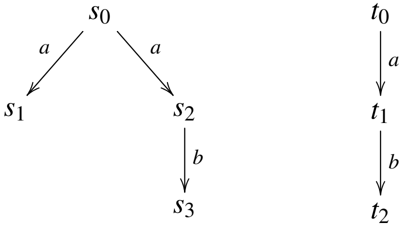
Larsen-Skou Bisimulation

- Let $\mathcal{S} = (S, i, \Sigma, \tau)$ be a labelled Markov process. An equivalence relation R on S is a **bisimulation** if whenever sRs', with $s, s' \in S$, we have that for all $a \in \mathcal{A}$ and every R-closed measurable set $A \in \Sigma$, $\tau_a(s, A) = \tau_a(s', A)$.
  Two states are bisimilar if they are related by a bisimulation relation.
- Can be extended to bisimulation between two different **LMP**s.

Logical Characterization

- $$\mathcal{L} ::== \mathsf{T} | \phi_1 \wedge \phi_2 | \langle a \rangle_q \phi$$
- We say $s \models \langle a \rangle_q \phi$ iff
$$\exists A \in \Sigma.(\forall s' \in A.s' \models \phi) \wedge (\tau_a(s, A) > q).$$
- Two systems are bisimilar iff they obey the same formulas of $\mathcal{L}$. [DEP 1998 LICS, I and C 2002]

That cannot be right?



Two processes that cannot be distinguished without negation.
The formula that distinguishes them is $\langle a\rangle(\neg\langle b\rangle\top)$.

But it is!

- $$\begin{tikzcd}
 & s_0 \arrow[dl, "{a[p]}"'] \arrow[dr, "{a[q]}"] & & & t_0 \arrow[d, "{a[r]}"] \\
s_1 & & s_2 \arrow[d, "b"] & & t_1 \arrow[d, "b"] \\
 & & s_3 & & t_2
\end{tikzcd}$$

  We add probabilities to the transitions.
- If $p + q < r$ or $p + q > r$ we can easily distinguish them.
- If $p + q = r$ and $p > 0$ then $q < r$ so $\langle a \rangle_r \langle b \rangle_1 \top$ distinguishes them.

Proof idea

- Show that the relation "s and s' satisfy exactly the same formulas" is a bisimulation.
- Can easily show that $\tau_a(s, A) = \tau_a(s', A)$ for A of the form $[\![\phi]\!]$.
- Use Dynkin's lemma to show that we get a well defined measure on the σ-algebra generated by such sets and the above equality holds.
- Use special properties of analytic spaces to show that this σ-algebra is the same as the original σ-algebra.

The Easy Direction

- Let R be a bisimulation relation on an LMP (S, Σ, τ_a). We prove by induction on ϕ that $\forall \phi \in \mathcal{L}$

$$\forall s, s' \in S. sRs' \Rightarrow s \models \phi \Leftrightarrow s' \models \phi.$$

- Base case trivial.
- $\wedge$ is obvious from Inductive Hypothesis.
- For $\phi = \langle a \rangle_q \psi$ we have that $[\![\psi]\!]$ is R-closed from inductive hypothesis. Thus

$$\tau_a(s, [\![\psi]\!]) = \tau_a(s', [\![\psi]\!])$$

and thus $sRs' \Rightarrow s \models \phi \Leftrightarrow s' \models \phi$.

Digression on Analytic Spaces

- An analytic set A is the image of a Polish space X (or a Borel subset of X) under a continuous (or measurable) function $f : X \to Y$, where Y is Polish. If (S, Σ) is a measurable space where S is an analytic set in some ambient topological space and Σ is the Borel σ-algebra on S.
- Analytic sets do not form a σ-algebra but they are in the completion of the Borel algebra under **any** measure. [Universally measurable.]
- Regular conditional probability densities can be defined on analytic spaces.

Amazing Facts about Analytic Spaces

- Given A an analytic space and $\sim$ an equivalence relation such that there is a *countable* family of real-valued measurable functions $f_i : S \longrightarrow \mathbf{R}$ such that

$$\forall s, s' \in S.s \sim s' \iff \forall f_i.f_i(s) = f_i(s')$$

  then the quotient space (Q, Ω) - where $Q = S/\sim$ and Ω is the finest σ-algebra making the canonical surjection $q : S \longrightarrow Q$ measurable - is also analytic.
- If an analytic space (S, Σ) has a sub-σ-algebra Σ_0 of Σ which separates points and is countably generated then Σ_0 is Σ! The Unique Structure Theorem (UST).

The big picture

1. We have LMP $(S, \Sigma, \mathsf{L}, \tau_a)$ and we want to quotient by $\simeq$ where $s \simeq s'$ if they agree on all formulas of the logic.

$$\begin{tikzcd} (S, \Sigma, \mathsf{L}, \tau_a) \arrow[d, "q"] \\ (S/\simeq, \Sigma/\simeq, \mathsf{L}, \rho_a) \end{tikzcd}$$

2. We want to define ρ_a in such a way that

$$\rho_a(q(s), B) = \tau_a(s, q^{-1}(B)).$$

3. Why?
4. In lieu of an answer: maps between LMP's satisfying the above condition are called "zigzags" and bisimulation can be defined as the existence of a span of zigzags.

ρ is well defined - I

- Easy to check that $q^{-1}(q(\llbracket\phi\rrbracket)) = \llbracket\phi\rrbracket$:
  $s \in q^{-1}(q(\llbracket\phi\rrbracket))$ implies that $q(s) \in q(\llbracket\phi\rrbracket)$, i.e. $\exists s' \in \llbracket\phi\rrbracket . s \simeq s'$, so $s \models \phi$ so $s \in \llbracket\phi\rrbracket$.
- Thus $q(\llbracket\phi\rrbracket)$ is measurable.
- Thus the σ-algebra generated -say, Λ - by $q(\llbracket\phi\rrbracket)$ is a sub-σ-algebra of Ω.
- Λ is countably generated and separates points so by UST it *is* Ω. Thus $q(\llbracket\phi\rrbracket)$ generates Ω.

ρ is well defined - II

- The collection $q([\![\phi]\!])$ is a π-system (because $\mathcal{L}_0$ has conjunction) and it generates Ω; thus if we can show that two measures agree on these sets they agree on all of Ω.
- If $q(s) = q(s') = t$ then $\tau_a(s, [\![\phi]\!]) = \tau_a(s', [\![\phi]\!])$ (simple interpolation).
- Thus $\tau_a(s, q^{-1}(q([\![\phi]\!]))) = \tau_a(s', q^{-1}(q([\![\phi]\!])))$ and hence ρ is well defined. We have $\rho_a(q(s), B) = \tau_a(s, q^{-1}(B))$.

Finishing the Argument

- Let X be any $\simeq$-closed subset of S.
- Then $q^{-1}(q(X)) = X$ and $q(X) \in \Omega$.
- If $s \simeq s'$ then $q(s) = q(s')$ and

$$\tau_a(s, X) = \tau_a(s, q^{-1}(q(X))) = \rho_a(q(s), q(X)) =$$

$$\rho_a(q(s'), q(X)) = \tau_a(s', q^{-1}(q(X))) = \tau_a(s', X).$$

Simulation

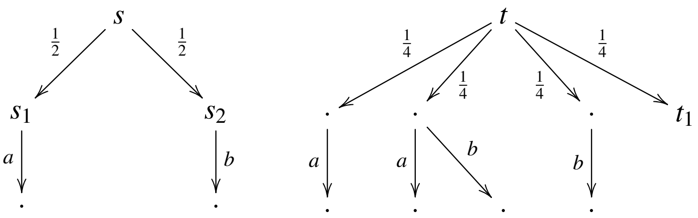
Let $\mathcal{S} = (S, \Sigma, \tau)$ be a labelled Markov process. A preorder R on S is a **simulation** if whenever sRs', we have that for all $a \in \mathcal{A}$ and every R-closed measurable set $A \in \Sigma$, $\tau_a(s, A) \leq \tau_a(s', A)$. We say s is simulated by s' if sRs' for some simulation relation R.

Logic for simulation?

- The logic used in the characterization has no negation, not even a limited negative construct.
- One can show that if s simulates s' then s satisfies all the formulas of $\mathcal{L}$ that s' satisfies.
- What about the converse?

Counter example!

In the following picture, t satisfies all formulas of $\mathcal{L}$ that s satisfies but t does not simulate s.



All transitions from s and t are labelled by a.

Counter example (contd.)

- A formula of $\mathcal{L}$ that is satisfied by t but not by s.

$$\langle a\rangle_0(\langle a\rangle_0\mathsf{T}\wedge\langle b\rangle_0\mathsf{T}).$$

- A formula with disjunction that is satisfied by s but not by t:

$$\langle a\rangle_{\frac{3}{4}}(\langle a\rangle_0\mathsf{T}\vee\langle b\rangle_0\mathsf{T}).$$

A logical characterization for simulation

- The logic $\mathcal{L}$ does **not** characterize simulation. One needs disjunction.

$$\mathcal{L}_\vee := \mathcal{L}\phi 1 \vee \phi_2.$$

- With this logic we have:
  An **LMP** s_1 simulates s_2 if and only if for every formula ϕ of $\mathcal{L}_\vee$ we have
  $$s_1 \models \phi \Rightarrow s_2 \models \phi.$$
- The only proof we know uses domain theory.

Other Logics

$$
\begin{array}{rcl}
\mathcal{L}_{\mathrm{Can}} & := & \mathcal{L}_0 \mid \mathrm{Can}(a) \\
\mathcal{L}_{\Delta} & := & \mathcal{L}_0 \mid \Delta_a \\
\mathcal{L}_{\neg} & := & \mathcal{L}_0 \mid \neg\phi \\
\mathcal{L}_{\vee} & := & \mathcal{L}_0 \mid \phi_1 \vee \phi_2 \\
\mathcal{L}_{\wedge} & := & \mathcal{L}_{\neg} \mid \bigwedge_{i \in \mathbf{N}} \phi_i
\end{array}
$$

where

$s \models \mathrm{Can}(a)$ to mean that $\tau_a(s, S) > 0$;
$s \models \Delta_a$ to mean that $\tau_a(s, S) = 0$.

We need $\mathcal{L}_{\vee}$ to characterise simulation.

Conclusions

- Strong probabilistic bisimulation is characterised by a very simple modal logic with no negative constructs.
- There is a logical characterisation of simulation.
- There is a “metric” on LMPs which is based on this logic.
- Why did the proof require so many subtle properties of analytic spaces? There is a more general definition of bisimulation for which the logical characterisation proof is “easy” but to prove that that definition coincides with this one in analytic spaces requires roughly the same proof as that given here.