Classification among Hidden Markov Models

S. Akshay - **Hugo Bazille** - Eric Fabre - Blaise Genest

18/06/2019

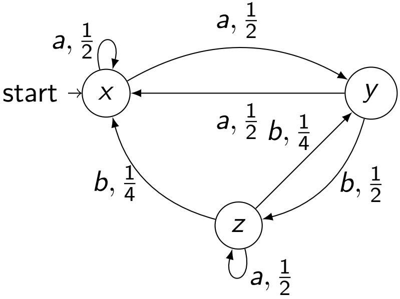
Summary

Introduction of the problem (1)

Framework

- Stochastic systems,
- Partial information.

$\Rightarrow$ Hidden Markov Models:



Introduction of the problem (2)

Classification

Given two systems A_1, A_2 and an observation w, decide which one produced it.

Encompasses diagnosis, opacity...

Introduction of the problem (2)

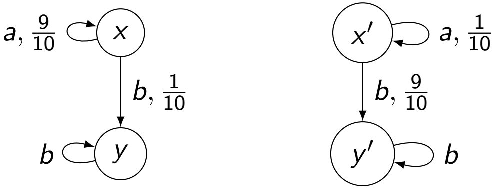
Classification

Given two systems A_1, A_2 and an observation w, decide which one produced it.

Encompasses diagnosis, opacity...

Can we classify...

- For sure?
- Almost sure?
- Limit sure?
- We cannot?



What is a classifier?

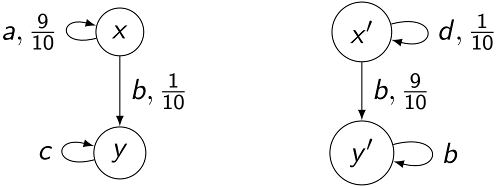
Function $f : \Sigma^* \to \{\perp, 1, 2\}$

What is a good classifier?

- Accurate?
- Reactive?
- No error?

Sure classification

Informally: ability to distinguish after some time.
Formally: $\forall w \in \Sigma^\infty, \exists v, w = vv', v \in L_1, v \notin L_2$.

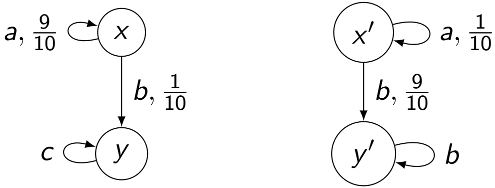


Theorem

Sure classification is decidable in PTIME, by deciding if $L_1^\infty \cap L_2^\infty = \emptyset$.

Almost sure classification

Informally: ability to distinguish after some time with probability 1.
Formally: $P(w \in \Sigma^\infty,\ \exists v, w = vv',\ v \in L_1,\ v \notin L_2) = 1$.

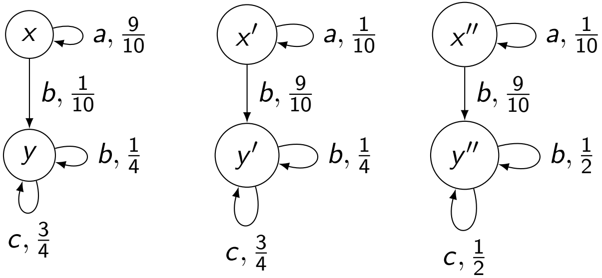


Theorem

Almost sure classification is PSPACE-complete, by deciding if $P(L_1^\infty \cap L_2^\infty) = 0$.

Limit sure classifiability

Informally: classify with arbitrarily high precision.
Formally: there is a classifier f that eventually answers correctly with probability $> 1 - \varepsilon$ for all $\varepsilon > 0$.



Main result on limit sure classifiability

Theorem

Limit sure classifiability is decidable in PTIME.

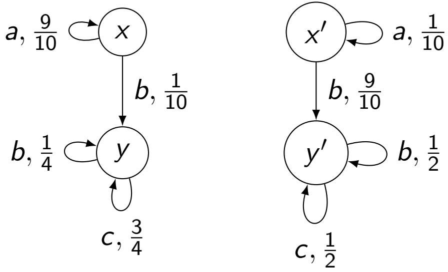
Main result on limit sure classifiability

Theorem

Limit sure classifiability is decidable in PTIME.

We want arbitrarily high precision:

- Transient components "do not matter",
- We mainly study BSCCs.



Study of BSCC

Two BSCCs are problematic if they:

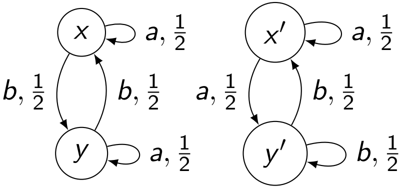
1. Are co-reachable,
2. Have the same stochastic language.

Study of BSCC

Two BSCCs are problematic if they:

1. Are co-reachable,
2. Have the same stochastic language.

How to check this? State by state?



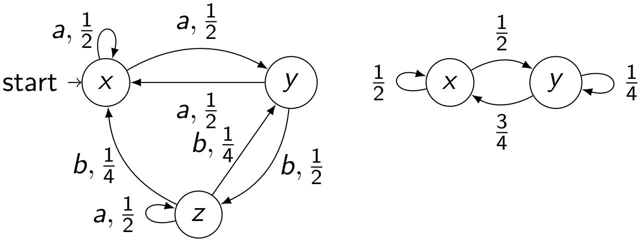
- $L_x \not\equiv L_{x'}$
- $L_x \not\equiv L_{y'}$
- ...

But $L_{\frac{1}{2}x+\frac{1}{2}y} \equiv L_{\frac{1}{2}x'+\frac{1}{2}y'}$.

How to be smart? (1)

Number of possible distributions: infinite!

Consider stationnary distributions σ_X on beliefs X.



For a belief X, σ_X is computable in PTIME.

Interest of stationary distributions

Theorem

The following are equivalent:

1. One cannot classify between A_1, A_2,
2. There exists an X in a BSCC of twin beliefs such that $(A_1, \sigma_X^1) \equiv (A_2, \sigma_X^2)$.

Interest of stationary distributions

Theorem

The following are equivalent:

1. One cannot classify between A_1, A_2,
2. There exists an X in a BSCC of twin beliefs such that $(A_1, \sigma_X^1) \equiv (A_2, \sigma_X^2)$.

One problem solved, but... still an exponential number of beliefs!

How to be smart? (2)

Have only a limited number of beliefs?
$A = A_1 \times A_2$, for a BSCC D_i of A and $(y_1, y_2) \in D_i$,

- $X_1 = \{x_1 \mid (x_1, y_2) \in D_i\}$,
- $X_2 = \{x_2 \mid (y_1, x_2) \in D_i\}$.

How to be smart? (2)

Have only a limited number of beliefs?
$A = A_1 \times A_2$, for a BSCC D_i of A and $(y_1, y_2) \in D_i$,

- $X_1 = \{x_1 \mid (x_1, y_2) \in D_i\}$,
- $X_2 = \{x_2 \mid (y_1, x_2) \in D_i\}$.

Theorem

NSC: check equivalence for such X_1, X_2.

- Polynomial number of such beliefs,
- Each check with Linear Programming: PTIME!

With tries?

User has a reset button:

- Can try again and again,
- Chooses the system randomly every time.

With tries?

User has a reset button:

- Can try again and again,
- Chooses the system randomly every time.

Attack-classifiability

Decide if there exists a reset strategy such that:

1. It will finish with probability 1,
2. It is limit sure classifiable after the last reset.

With tries?

User has a reset button:

- Can try again and again,
- Chooses the system randomly every time.

Attack-classifiability

Decide if there exists a reset strategy such that:

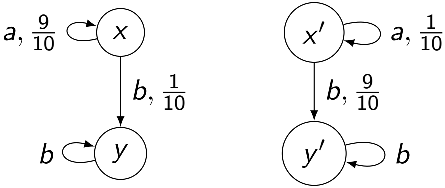
1. It will finish with probability 1,
2. It is limit sure classifiable after the last reset.

$1-\varepsilon$ attacker-classifiability

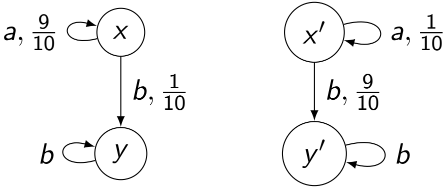
With ε fixed, decide if there exists a reset strategy such that:

1. It will finish with probability 1,
2. Classification will be correct with probability $1-\varepsilon$ after last reset.

An example



An example



- Not attack classifiable,
- $\forall \varepsilon$, $1 - \varepsilon$ attack classifiable.

Results on these variants

Theorem

- Attack-classifiability is PSPACE-complete.
- $1 - \varepsilon$ attacker-classifiability is undecidable.

Results on these variants

Theorem

- Attack-classifiability is PSPACE-complete.
- $1 - \varepsilon$ attacker-classifiability is undecidable.

Idea of proofs:

Attack-classifiability:

- Find subpart of the systems that are classifiable,
- Hardness: reduction from language inclusion for finite automata.

$1 - \varepsilon$ attacker-classifiability:

- Reduction from 0 and 1 isolation problem for PFA.

Summary

Classifiabilities

1. Sure: a word in only one language.
2. Almost Sure: a word in only one language with probability 1.
3. Limit Sure: probability of error decreases to 0.
4. Attack: Limit Sure with tries.
5. $1-\varepsilon$ Attack: decide with a fixed threshold of error with tries.

Class	Sure	Almost Sure	Limit Sure	Attack	$1-\varepsilon$ attack
Cplxt	PTIME	PSPACE	PTIME	PSPACE	undecidable

Strong links with:

- Distance 1 problem: determine if

$$sup_{W \in \Sigma^{\infty}} |P_1(W) - P_2(W)| = 1$$

- AFF-diagnosability and ε-diagnosability.

Strong links with:

- Distance 1 problem: determine if

$$sup_{W \in \Sigma^\infty} |P_1(W) - P_2(W)| = 1$$

- AFF-diagnosability and ε-diagnosability.

Questions time!