

Automata-based stream processing

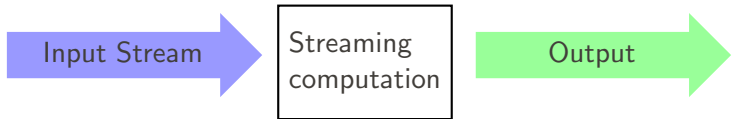
ICALP 2017

Rajeev Alur, Konstantinos Mamouras, and Caleb Stanford

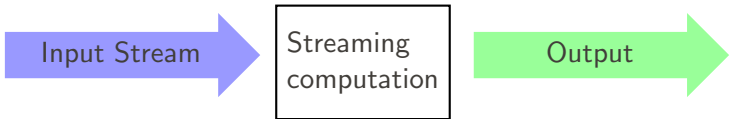
University of Pennsylvania

July 10, 2017

Streaming model



Streaming model

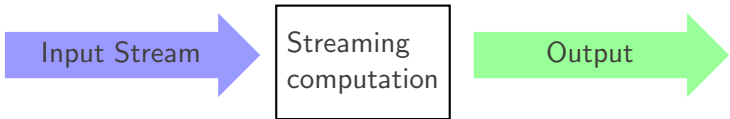


..., (clickid₂, x₂, y₂, data₂), (clickid₁, x₁, y₁, data₁)

..., (IPaddr₂, request₂), (IPaddr₁, request₁)

..., BP₄, BP₃, BP₂, BP₁

Streaming model



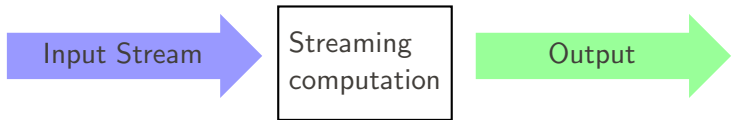
..., (clickid₂, x₂, y₂, data₂), (clickid₁, x₁, y₁, data₁)

..., (IPaddr₂, request₂), (IPaddr₁, request₁)

..., BP₄, BP₃, BP₂, BP₁

- Compute *quantitative* property of input stream
- Output in real-time
- Efficient space
- Efficient time per element

Streaming model



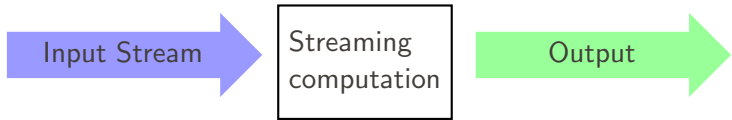
“Maximum average during an episode”

128, 77, 130, 75, 147, 87, 149, 85, 139, 80, 152, 89, 146, 84

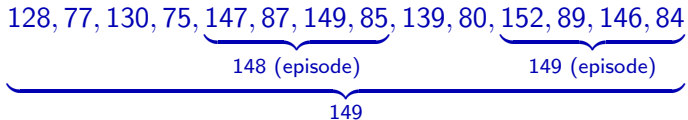
148 (episode) 149 (episode)

149

Streaming model



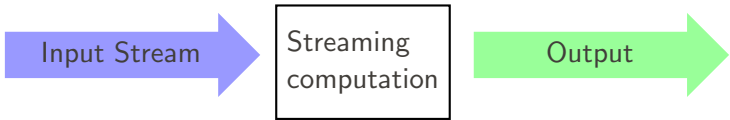
“Maximum average during an episode”



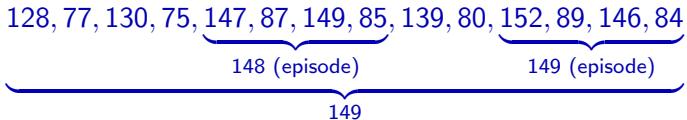
Quantitative Regular Expressions (PLDI 2017):

```
iter(split( iter((x < 140)  $\mapsto$  0, 0, +),
  combine( iter((x > 140)  $\mapsto$  x, 0, +),
    iter((x > 140)  $\mapsto$  1, 0, +), /) ), 0, max)
```

Streaming model



“Maximum average during an episode”



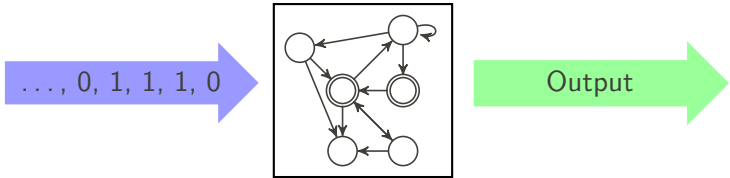
Quantitative Regular Expressions (PLDI 2017):

```
iter(split( iter((x < 140) ↦ 0, 0, +),
           combine( iter((x > 140) ↦ x, 0, +),
                   iter((x > 140) ↦ 1, 0, +), /) ), 0, max)
```

Can automata do this?

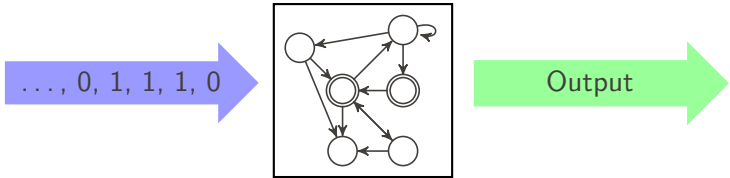
Why automata?

Why automata?



- Efficient space ✓
 - Efficient time per element ✓
 - NFAs > DFAs
- } "Streamability"

Why automata?



- Efficient space ✓
 - Efficient time per element ✓
 - NFAs > DFAs
- } "Streamability"
- Expressiveness ✗
 - Succinctness ✗

Our model

Stream Automata:

Weighted Automata + Nesting + Parallelism + Typing Restrictions
Nondeterminism (Unambiguity, Parallel Consistency)

Unambiguous Nondet. + Nesting + Consistent Parallelism ✓ (Thm. 4)

Our model

Stream Automata:

$\underbrace{\text{Weighted Automata}}_{\text{Nondeterminism}} + \text{Nesting} + \text{Parallelism} + \underbrace{\text{Typing Restrictions}}_{\substack{\text{(Unambiguity,} \\ \text{Parallel Consistency)}}$

Unambiguous Nondet. + Nesting + Consistent Parallelism ✓ (Thm. 4)

Nondeterminism + Nesting + Consistent Parallelism ✗ (Thm. 5)

Unambiguous Nondeterminism + Nesting + Parallelism ✗ (Thm. 6)

Our model

Stream Automata:

$\underbrace{\text{Weighted Automata}}_{\text{Nondeterminism}} + \text{Nesting} + \text{Parallelism} + \underbrace{\text{Typing Restrictions}}_{\substack{\text{(Unambiguity,} \\ \text{Parallel Consistency)}}$

Unambiguous Nondet. + Nesting + Consistent Parallelism ✓ (Thm. 4)

Nondeterminism + Nesting + Consistent Parallelism ✗ (Thm. 5)

Unambiguous Nondeterminism + Nesting + Parallelism ✗ (Thm. 6)

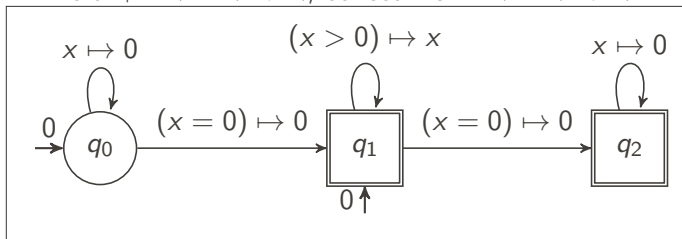
Outline:

- Features — Nondeterminism, Nesting, Parallelism
- Nondeterminism + Nesting ✗
- Unambiguous Nondeterminism + Nesting ✓
- Future work

Nondeterministic (symbolic) weighted automata

\mathcal{M}_1 computing MAXBLOCKSUM

fold $+$: $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$, collect \max : $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$

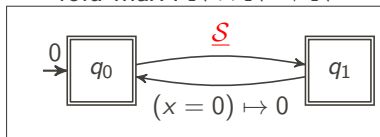


	3	1	0	3	0	2	3	
q_0	q_0	q_0	q_0	q_0	q_1	q_1	q_1	
	0	0	0	0	0	2	5	$\implies 5$
q_0	q_0	q_0	q_1	q_1	q_2	q_2	q_2	$\implies 5$
	0	0	0	3	3	3	3	$\implies 3$
q_1	q_1	q_1	q_2	q_2	q_2	q_2	q_2	
	0	3	4	4	4	4	4	$\implies 4$

Nesting

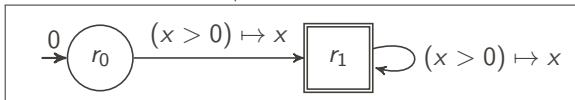
\mathcal{M}_2 computing MAXBLOCKSUM

fold $\max : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$



\mathcal{S} computing SUM

fold $+$: $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$

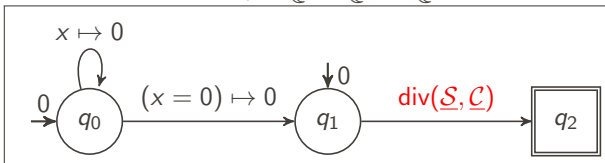


	3	1	0	3	0	2	3	
	q_0	q_1	q_0	q_1	q_0	q_1		
	0	4	4	4	4	5	\Rightarrow	5
	\mathcal{S}		\mathcal{S}		\mathcal{S}			
	(0 3 4)		(0 3)		(0 2 5)			

Parallelism

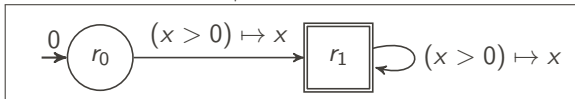
\mathcal{M}_3 computing LASTBLOCKAVERAGE

fold $+$: $\mathbb{Q} \times \mathbb{Q} \rightarrow \mathbb{Q}$



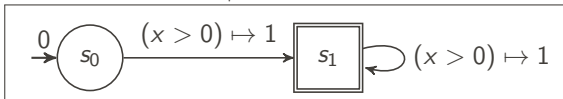
$\underline{\mathcal{S}}$ computing SUM

fold $+$: $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$



$\underline{\mathcal{C}}$ computing COUNT

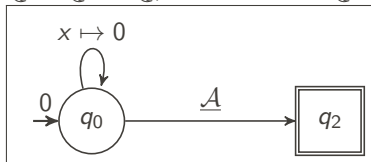
fold $+$: $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$



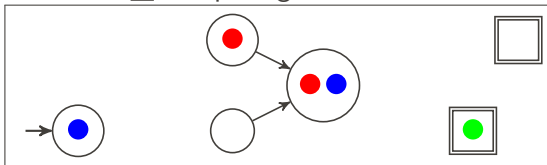
Nondeterminism + Nesting ~~X~~

\mathcal{M}_4 computing `MINSUFFIXAVERAGE`

fold $+$: $\mathbb{Q} \times \mathbb{Q} \rightarrow \mathbb{Q}$, collect min : $\mathbb{Q} \times \mathbb{Q} \rightarrow \mathbb{Q}$



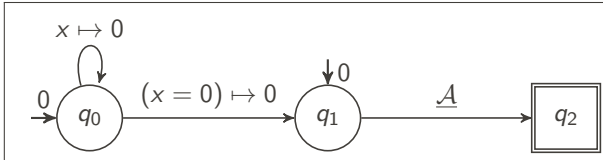
\underline{A} computing `AVERAGE`



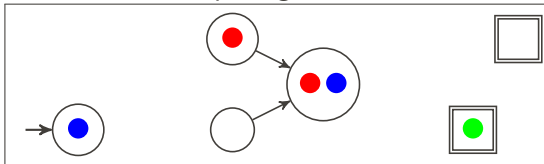
Unambiguous Nondeterminism + Nesting ✓

\mathcal{M}_5 computing LASTBLOCKAVERAGE

fold $+$: $\mathbb{Q} \times \mathbb{Q} \rightarrow \mathbb{Q}$



$\underline{\mathcal{A}}$ computing AVERAGE



Future work

Expressiveness — equivalence of models

“Flat” model — more immediate evaluation algorithm

Optimization



Other models



Nested weighted automata

Register automata

Expressiveness

Output value \longrightarrow computed term

“MSO-definable string to tree transformations”

Capturing streamability