

Automata Learning for Streaming Data

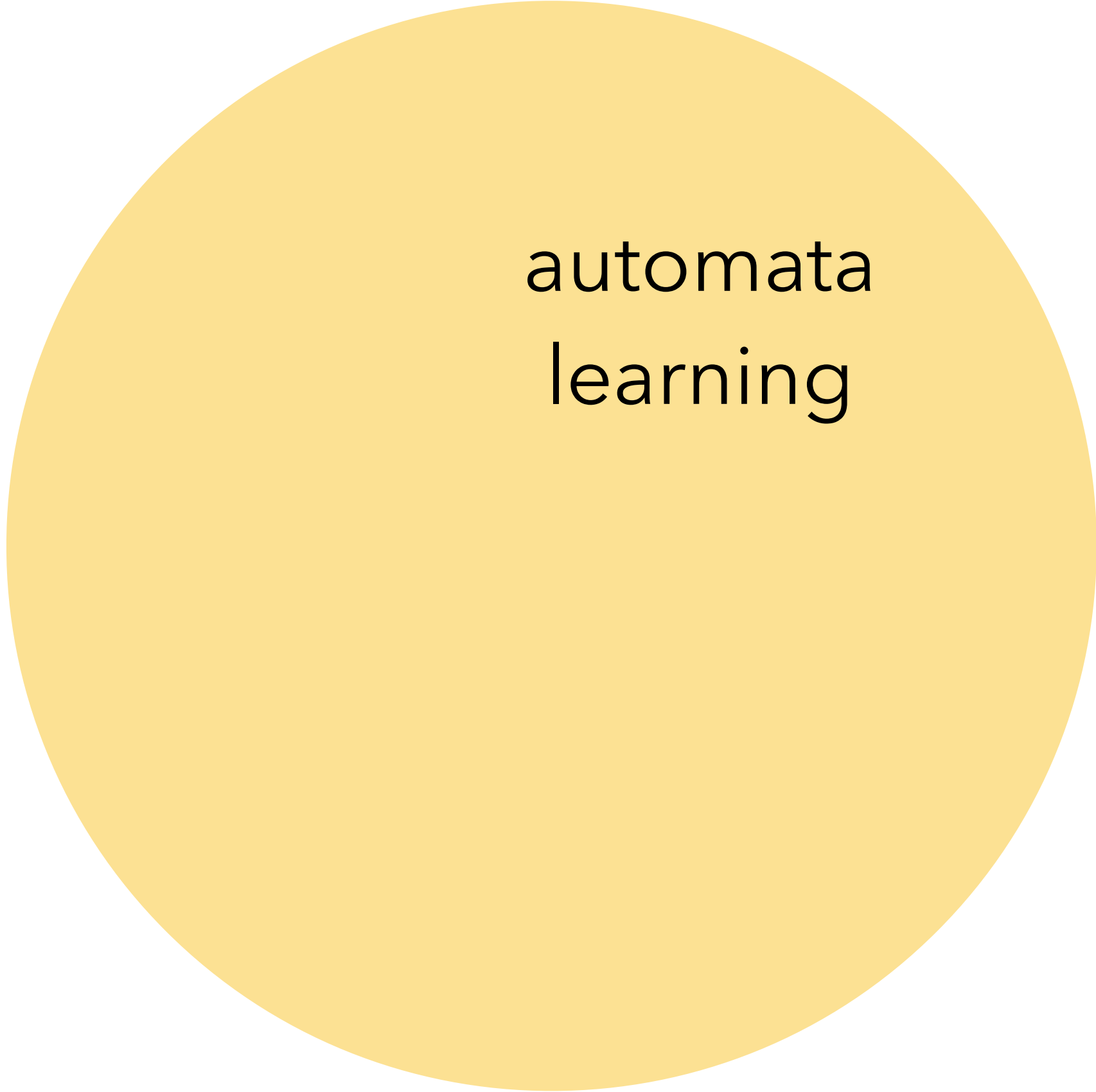
Benedikt Bollig

CNRS, Université Paris-Saclay, ENS Paris-Saclay, LMF

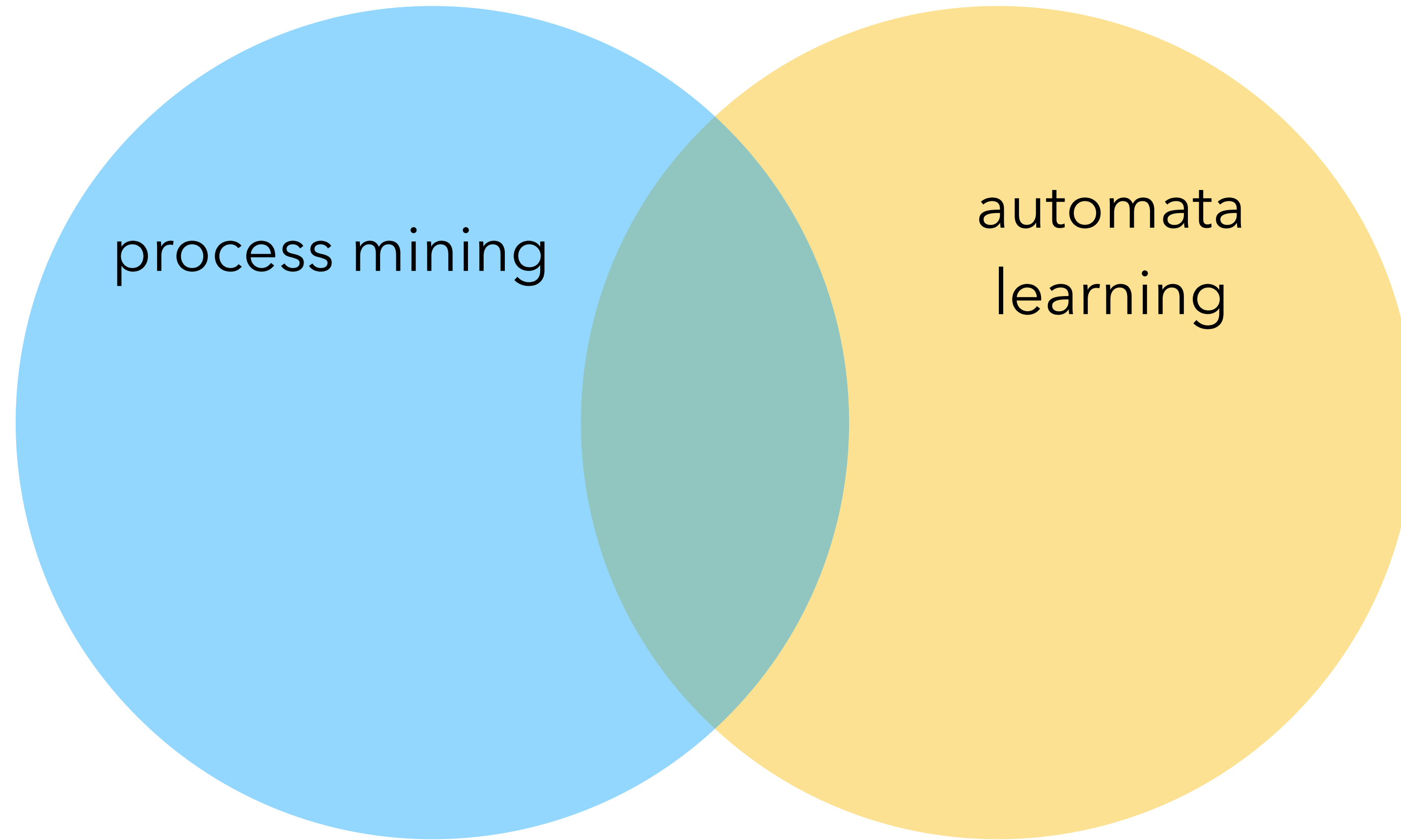
joint work with Matthias Függer and Thomas Nowak

Journées Synthèse de Programmes

LaBRI, Bordeaux, November 26-27, 2024

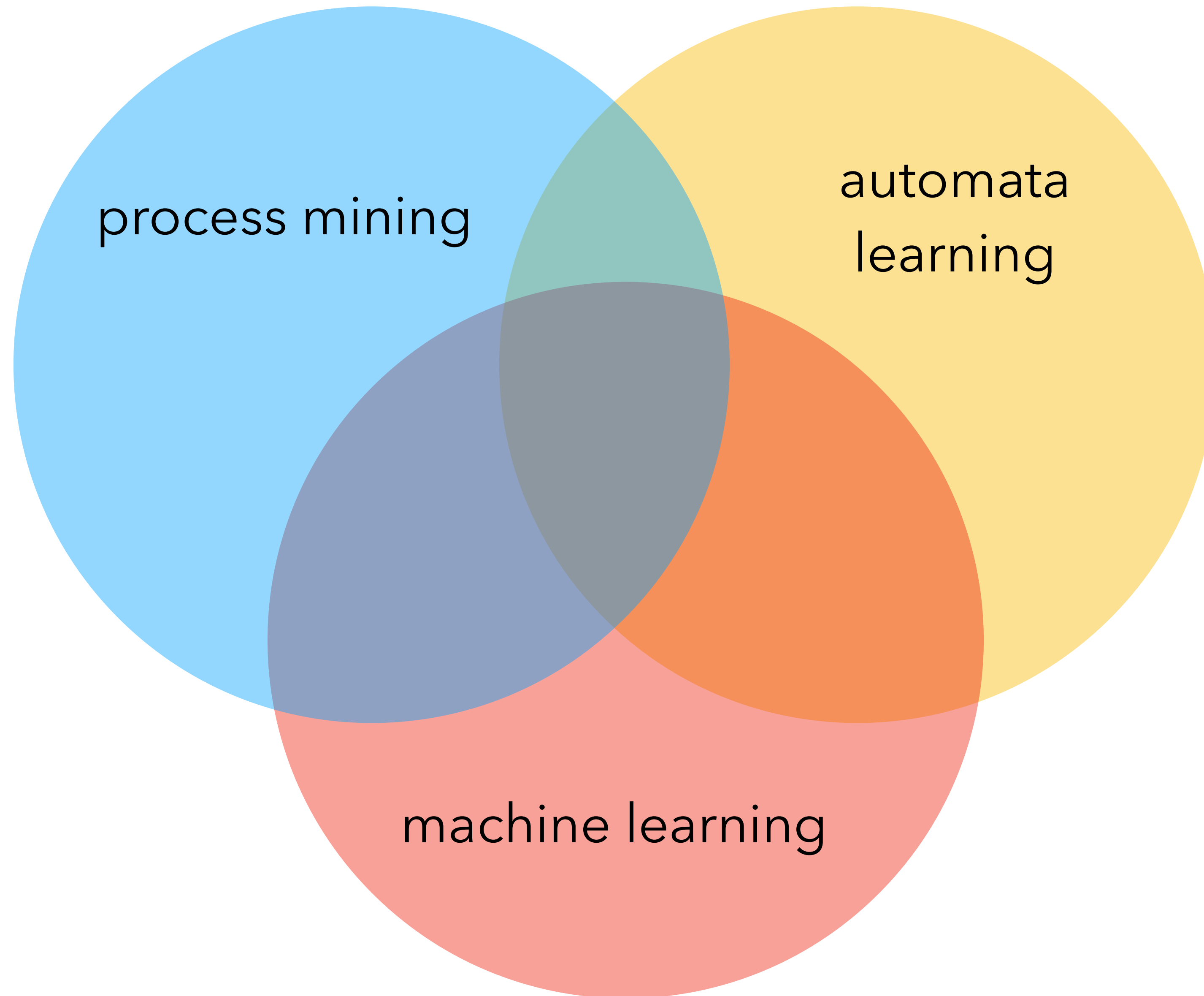


automata
learning



process mining

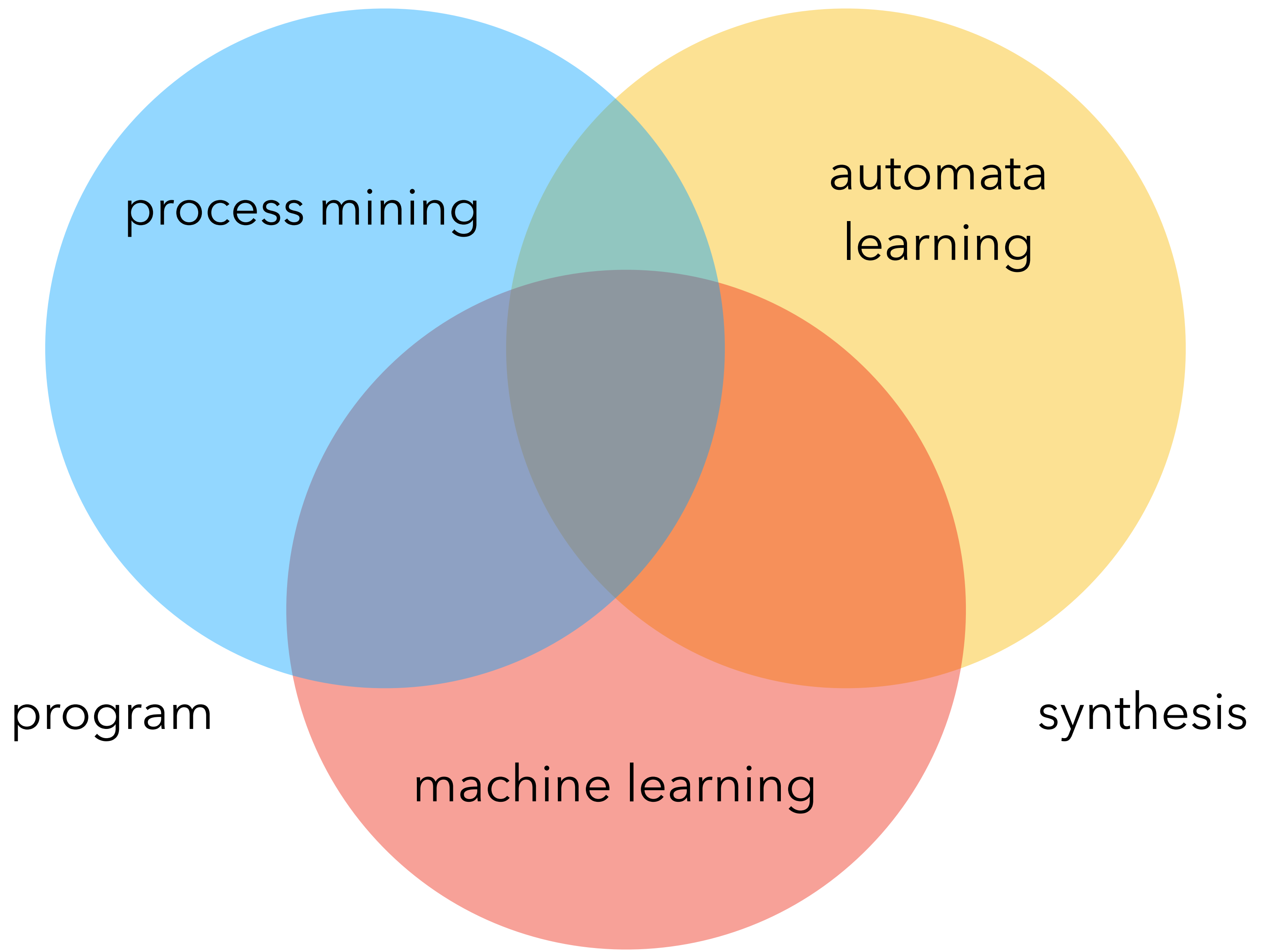
automata
learning



process mining

automata
learning

machine learning



process mining

automata
learning

program

synthesis

machine learning

Event-log prediction

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Felix Mannhardt & Daan Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

case:concept:name	concept:name	time:timestamp	Age	LacticAcid	Leucocytes
XJ	ER Registration	2013-11-07 08:18:29+00:00	90.0		
XJ	ER Triage	2013-11-07 08:29:18+00:00			
XJ	ER Sepsis Triage	2013-11-07 08:37:32+00:00			
XJ	LacticAcid	2013-11-07 08:51:00+00:00		1.4	
XJ	Leucocytes	2013-11-07 08:51:00+00:00			296.2
XJ	CRP	2013-11-07 08:51:00+00:00			
XJ	IV Liquid	2013-11-07 09:05:57+00:00			
XJ	IV Antibiotics	2013-11-07 10:05:58+00:00			
XJ	Admission NC	2013-11-07 11:11:34+00:00			
XJ	Leucocytes	2013-11-08 08:00:00+00:00			297.6
I	ER Registration	2013-11-09 09:21:03+00:00	80.0		
I	ER Triage	2013-11-09 09:34:28+00:00			
I	ER Sepsis Triage	2013-11-09 09:34:41+00:00			
I	IV Antibiotics	2013-11-09 09:34:56+00:00			
I	LacticAcid	2013-11-09 09:42:00+00:00		1.1	
I	CRP	2013-11-09 09:42:00+00:00			
I	Leucocytes	2013-11-09 09:42:00+00:00			11.0
I	Admission NC	2013-11-09 10:56:59+00:00			
WEA	ER Registration	2013-11-09 12:04:54+00:00	70.0		
WEA	ER Triage	2013-11-09 12:18:39+00:00			
WEA	ER Sepsis Triage	2013-11-09 12:18:59+00:00			
WEA	CRP	2013-11-09 12:40:00+00:00			
WEA	Leucocytes	2013-11-09 12:40:00+00:00			12.8
WEA	IV Liquid	2013-11-09 15:04:35+00:00			
WEA	IV Antibiotics	2013-11-09 15:04:41+00:00			
I	CRP	2013-11-11 08:00:00+00:00			
I	Leucocytes	2013-11-11 08:00:00+00:00			8.7
OT	ER Registration	2013-11-11 12:48:21+00:00	75.0		
OT	ER Triage	2013-11-11 13:16:57+00:00			
OT	ER Sepsis Triage	2013-11-11 13:17:29+00:00			

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

Felix Mannhardt & Daan Blinde (2018):
[Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

case id
↙

case:concept:name	concept:name	time:timestamp	Age	LacticAcid	Leucocytes
XJ	ER Registration	2013-11-07 08:18:29+00:00	90.0		
XJ	ER Triage	2013-11-07 08:29:18+00:00			
XJ	ER Sepsis Triage	2013-11-07 08:37:32+00:00			
XJ	LacticAcid	2013-11-07 08:51:00+00:00		1.4	
XJ	Leucocytes	2013-11-07 08:51:00+00:00			296.2
XJ	CRP	2013-11-07 08:51:00+00:00			
XJ	IV Liquid	2013-11-07 09:05:57+00:00			
XJ	IV Antibiotics	2013-11-07 10:05:58+00:00			
XJ	Admission NC	2013-11-07 11:11:34+00:00			
XJ	Leucocytes	2013-11-08 08:00:00+00:00			297.6
I	ER Registration	2013-11-09 09:21:03+00:00	80.0		
I	ER Triage	2013-11-09 09:34:28+00:00			
I	ER Sepsis Triage	2013-11-09 09:34:41+00:00			
I	IV Antibiotics	2013-11-09 09:34:56+00:00			
I	LacticAcid	2013-11-09 09:42:00+00:00		1.1	
I	CRP	2013-11-09 09:42:00+00:00			
I	Leucocytes	2013-11-09 09:42:00+00:00			11.0
I	Admission NC	2013-11-09 10:56:59+00:00			
WEA	ER Registration	2013-11-09 12:04:54+00:00	70.0		
WEA	ER Triage	2013-11-09 12:18:39+00:00			
WEA	ER Sepsis Triage	2013-11-09 12:18:59+00:00			
WEA	CRP	2013-11-09 12:40:00+00:00			
WEA	Leucocytes	2013-11-09 12:40:00+00:00			12.8
WEA	IV Liquid	2013-11-09 15:04:35+00:00			
WEA	IV Antibiotics	2013-11-09 15:04:41+00:00			
I	CRP	2013-11-11 08:00:00+00:00			
I	Leucocytes	2013-11-11 08:00:00+00:00			8.7
OT	ER Registration	2013-11-11 12:48:21+00:00	75.0		
OT	ER Triage	2013-11-11 13:16:57+00:00			
OT	ER Sepsis Triage	2013-11-11 13:17:29+00:00			

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Felix Mannhardt & Daan Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

case:concept:name	concept:name	time:timestamp	Age	LacticAcid	Leucocytes
XJ	ER Registration	2013-11-07 08:18:29+00:00	90.0		
XJ	ER Triage	2013-11-07 08:29:18+00:00			
XJ	ER Sepsis Triage	2013-11-07 08:37:32+00:00			
XJ	LacticAcid	2013-11-07 08:51:00+00:00		1.4	
XJ	Leucocytes	2013-11-07 08:51:00+00:00			296.2
XJ	CRP	2013-11-07 08:51:00+00:00			
XJ	IV Liquid	2013-11-07 09:05:57+00:00			
XJ	IV Antibiotics	2013-11-07 10:05:58+00:00			
XJ	Admission NC	2013-11-07 11:11:34+00:00			
XJ	Leucocytes	2013-11-08 08:00:00+00:00			297.6
I	ER Registration	2013-11-09 09:21:03+00:00	80.0		
I	ER Triage	2013-11-09 09:34:28+00:00			
I	ER Sepsis Triage	2013-11-09 09:34:41+00:00			
I	IV Antibiotics	2013-11-09 09:34:56+00:00			
I	LacticAcid	2013-11-09 09:42:00+00:00		1.1	
I	CRP	2013-11-09 09:42:00+00:00			
I	Leucocytes	2013-11-09 09:42:00+00:00			11.0
I	Admission NC	2013-11-09 10:56:59+00:00			
WEA	ER Registration	2013-11-09 12:04:54+00:00	70.0		
WEA	ER Triage	2013-11-09 12:18:39+00:00			
WEA	ER Sepsis Triage	2013-11-09 12:18:59+00:00			
WEA	CRP	2013-11-09 12:40:00+00:00			
WEA	Leucocytes	2013-11-09 12:40:00+00:00			12.8
WEA	IV Liquid	2013-11-09 15:04:35+00:00			
WEA	IV Antibiotics	2013-11-09 15:04:41+00:00			
I	CRP	2013-11-11 08:00:00+00:00			
I	Leucocytes	2013-11-11 08:00:00+00:00			8.7
OT	ER Registration	2013-11-11 12:48:21+00:00	75.0		
OT	ER Triage	2013-11-11 13:16:57+00:00			
OT	ER Sepsis Triage	2013-11-11 13:17:29+00:00			

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

Felix Mannhardt & Daan Blinde (2018):
[Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

case:concept:name	concept:name	time:timestamp	Age	LacticAcid	Leucocytes
XJ	ER Registration	2013-11-07 08:18:29+00:00	90.0		
XJ	ER Triage	2013-11-07 08:29:18+00:00			
XJ	ER Sepsis Triage	2013-11-07 08:37:32+00:00			
XJ	LacticAcid	2013-11-07 08:51:00+00:00		1.4	
XJ	Leucocytes	2013-11-07 08:51:00+00:00			296.2
XJ	CRP	2013-11-07 08:51:00+00:00			
XJ	IV Liquid	2013-11-07 09:05:57+00:00			
XJ	IV Antibiotics	2013-11-07 10:05:58+00:00			
XJ	Admission NC	2013-11-07 11:11:34+00:00			
XJ	Leucocytes	2013-11-08 08:00:00+00:00			297.6
I	ER Registration	2013-11-09 09:21:03+00:00	80.0		
I	ER Triage	2013-11-09 09:34:28+00:00			
I	ER Sepsis Triage	2013-11-09 09:34:41+00:00			
I	IV Antibiotics	2013-11-09 09:34:56+00:00			
I	LacticAcid	2013-11-09 09:42:00+00:00		1.1	
I	CRP	2013-11-09 09:42:00+00:00			
I	Leucocytes	2013-11-09 09:42:00+00:00			11.0
I	Admission NC	2013-11-09 10:56:59+00:00			
WEA	ER Registration	2013-11-09 12:04:54+00:00	70.0		
WEA	ER Triage	2013-11-09 12:18:39+00:00			
WEA	ER Sepsis Triage	2013-11-09 12:18:59+00:00			
WEA	CRP	2013-11-09 12:40:00+00:00			
WEA	Leucocytes	2013-11-09 12:40:00+00:00			12.8
WEA	IV Liquid	2013-11-09 15:04:35+00:00			
WEA	IV Antibiotics	2013-11-09 15:04:41+00:00			
I	CRP	2013-11-11 08:00:00+00:00			
I	Leucocytes	2013-11-11 08:00:00+00:00			8.7
OT	ER Registration	2013-11-11 12:48:21+00:00	75.0		
OT	ER Triage	2013-11-11 13:16:57+00:00			
OT	ER Sepsis Triage	2013-11-11 13:17:29+00:00			

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Mannhardt & Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

event →

case id	activity/action				
case:concept:name	concept:name	time:timestamp	Age	LacticAcid	Leucocytes
XJ	ER Registration	2013-11-07 08:18:29+00:00	90.0		
XJ	ER Triage	2013-11-07 08:29:18+00:00			
XJ	ER Sepsis Triage	2013-11-07 08:37:32+00:00			
XJ	LacticAcid	2013-11-07 08:51:00+00:00		1.4	
XJ	Leucocytes	2013-11-07 08:51:00+00:00			296.2
XJ	CRP	2013-11-07 08:51:00+00:00			
XJ	IV Liquid	2013-11-07 09:05:57+00:00			
XJ	IV Antibiotics	2013-11-07 10:05:58+00:00			
XJ	Admission NC	2013-11-07 11:11:34+00:00			
XJ	Leucocytes	2013-11-08 08:00:00+00:00			297.6
I	ER Registration	2013-11-09 09:21:03+00:00	80.0		
I	ER Triage	2013-11-09 09:34:28+00:00			
I	ER Sepsis Triage	2013-11-09 09:34:41+00:00			
I	IV Antibiotics	2013-11-09 09:34:56+00:00			
I	LacticAcid	2013-11-09 09:42:00+00:00		1.1	
I	CRP	2013-11-09 09:42:00+00:00			
I	Leucocytes	2013-11-09 09:42:00+00:00			11.0
I	Admission NC	2013-11-09 10:56:59+00:00			
WEA	ER Registration	2013-11-09 12:04:54+00:00	70.0		
WEA	ER Triage	2013-11-09 12:18:39+00:00			
WEA	ER Sepsis Triage	2013-11-09 12:18:59+00:00			
WEA	CRP	2013-11-09 12:40:00+00:00			
WEA	Leucocytes	2013-11-09 12:40:00+00:00			12.8
WEA	IV Liquid	2013-11-09 15:04:35+00:00			
WEA	IV Antibiotics	2013-11-09 15:04:41+00:00			
I	CRP	2013-11-11 08:00:00+00:00			
I	Leucocytes	2013-11-11 08:00:00+00:00			8.7
OT	ER Registration	2013-11-11 12:48:21+00:00	75.0		
OT	ER Triage	2013-11-11 13:16:57+00:00			
OT	ER Sepsis Triage	2013-11-11 13:17:29+00:00			

historical
event data

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Mannhardt & Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

case id	activity/action				
case:concept:name	concept:name	time:timestamp	Age	LacticAcid	Leucocytes
XJ	ER Registration	2013-11-07 08:18:29+00:00	90.0		
XJ	ER Triage	2013-11-07 08:29:18+00:00			
XJ	ER Sepsis Triage	2013-11-07 08:37:32+00:00			
XJ	LacticAcid	2013-11-07 08:51:00+00:00		1.4	
XJ	Leucocytes	2013-11-07 08:51:00+00:00			296.2
XJ	CRP	2013-11-07 08:51:00+00:00			
XJ	IV Liquid	2013-11-07 09:05:57+00:00			
XJ	IV Antibiotics	2013-11-07 10:05:58+00:00			
XJ	Admission NC	2013-11-07 11:11:34+00:00			
XJ	Leucocytes	2013-11-08 08:00:00+00:00			297.6
I	ER Registration	2013-11-09 09:21:03+00:00	80.0		
I	ER Triage	2013-11-09 09:34:28+00:00			
I	ER Sepsis Triage	2013-11-09 09:34:41+00:00			
I	IV Antibiotics	2013-11-09 09:34:56+00:00			
I	LacticAcid	2013-11-09 09:42:00+00:00		1.1	
I	CRP	2013-11-09 09:42:00+00:00			
I	Leucocytes	2013-11-09 09:42:00+00:00			11.0
I	Admission NC	2013-11-09 10:56:59+00:00			
WEA	ER Registration	2013-11-09 12:04:54+00:00	70.0		
WEA	ER Triage	2013-11-09 12:18:39+00:00			
WEA	ER Sepsis Triage	2013-11-09 12:18:59+00:00			
WEA	CRP	2013-11-09 12:40:00+00:00			
WEA	Leucocytes	2013-11-09 12:40:00+00:00			12.8
WEA	IV Liquid	2013-11-09 15:04:35+00:00			
?	?	?	?	?	?

historical event data

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Mannhardt & Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

case id →

event →

case:concept:name	concept:name	time:timestamp	Age	LacticAcid	Leucocytes
XJ	ER Registration	2013-11-07 08:18:29+00:00	90.0		
XJ	ER Triage	2013-11-07 08:29:18+00:00			
XJ	ER Sepsis Triage	2013-11-07 08:37:32+00:00			
XJ	LacticAcid	2013-11-07 08:51:00+00:00		1.4	
XJ	Leucocytes	2013-11-07 08:51:00+00:00			296.2
XJ	CRP	2013-11-07 08:51:00+00:00			
XJ	IV Liquid	2013-11-07 09:05:57+00:00			
XJ	IV Antibiotics	2013-11-07 10:05:58+00:00			
XJ	Admission NC	2013-11-07 11:11:34+00:00			
XJ	Leucocytes	2013-11-08 08:00:00+00:00			297.6
I	ER Registration	2013-11-09 09:21:03+00:00	80.0		
I	ER Triage	2013-11-09 09:34:28+00:00			
I	ER Sepsis Triage	2013-11-09 09:34:41+00:00			
I	IV Antibiotics	2013-11-09 09:34:56+00:00			
I	LacticAcid	2013-11-09 09:42:00+00:00		1.1	
I	CRP	2013-11-09 09:42:00+00:00			
I	Leucocytes	2013-11-09 09:42:00+00:00			11.0
I	Admission NC	2013-11-09 10:56:59+00:00			
WEA	ER Registration	2013-11-09 12:04:54+00:00	70.0		
WEA	ER Triage	2013-11-09 12:18:39+00:00			
WEA	ER Sepsis Triage	2013-11-09 12:18:59+00:00			
WEA	CRP	2013-11-09 12:40:00+00:00			
WEA	Leucocytes	2013-11-09 12:40:00+00:00			12.8
WEA	IV Liquid	2013-11-09 15:04:35+00:00			
WEA	?	?	?	?	?

historical event data

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Mannhardt & Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

case id	activity/action
case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	?

event

historical event data

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

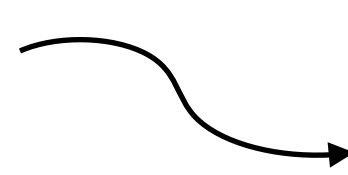
[Mannhardt & Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

case id activity/action $\in \Sigma$ finite alphabet

event 

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	?

historical event data

case id 

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Mannhardt & Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

event 

case id activity/action $\in \Sigma$ finite alphabet

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage
OT	ER Sepsis Triage

- resource allocation
- risk management
- predict energy consumption

[Guerard, Pousseur, Rivoire \(2020\): Inférence Grammaticale pour la Prévion de la Consommation Énergétique](#)

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Mannhardt & Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

event 

case id activity/action $\in \Sigma$ finite alphabet

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage
OT	ER Sepsis Triage

- resource allocation
- risk management
- predict energy consumption

[Guerard, Pousseur, Rivoire \(2020\): Inférence Grammaticale pour la Prévision de la Consommation Énergétique](#)

batch vs. streaming

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Mannhardt & Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

case id activity/action $\in \Sigma$ finite alphabet

event \rightarrow

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage
OT	ER Sepsis Triage

- resource allocation
- risk management
- predict energy consumption

[Guerard, Pousseur, Rivoire \(2020\): Inférence Grammaticale pour la Prédiction de la Consommation Énergétique](#)

model trained offline on different dataset

\uparrow
batch vs. streaming

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Mannhardt & Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

event

case id activity/action $\in \Sigma$ finite alphabet

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage
OT	ER Sepsis Triage

- resource allocation
- risk management
- predict energy consumption

[Guerard, Pousseur, Rivoire \(2020\): Inférence Grammaticale pour la Prévision de la Consommation Énergétique](#)

model trained offline on different dataset

batch vs. streaming

model trained online while observing and predicting events

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Mannhardt & Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

case id activity/action $\in \Sigma$ finite alphabet

event \rightarrow

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage
OT	ER Sepsis Triage

Event-log prediction

Real-world dataset:

Trajectories of 1050 patients with symptoms of a sepsis condition in a Dutch hospital, from their registration in the emergency room until their discharge.

[Mannhardt & Blinde \(2018\): Analyzing the Trajectories of Patients with Sepsis using Process Mining.](#)

case id activity/action $\in \Sigma$ finite alphabet

event \rightarrow

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage
OT	ER Sepsis Triage

Number of actions: 16

Number of cases: 1050

Average length of case: 14.49

Number of events: 15214

Action occurrences:

ER Registration	1050
ER Triage	1053
ER Sepsis Triage	1049
LacticAcid	1466
Leucocytes	3383
CRP	3262
IV Liquid	753
IV Antibiotics	823
Admission NC	1182
Release A	671
Return ER	294
Admission IC	117
Release B	56
Release C	25
Release D	24
Release E	6

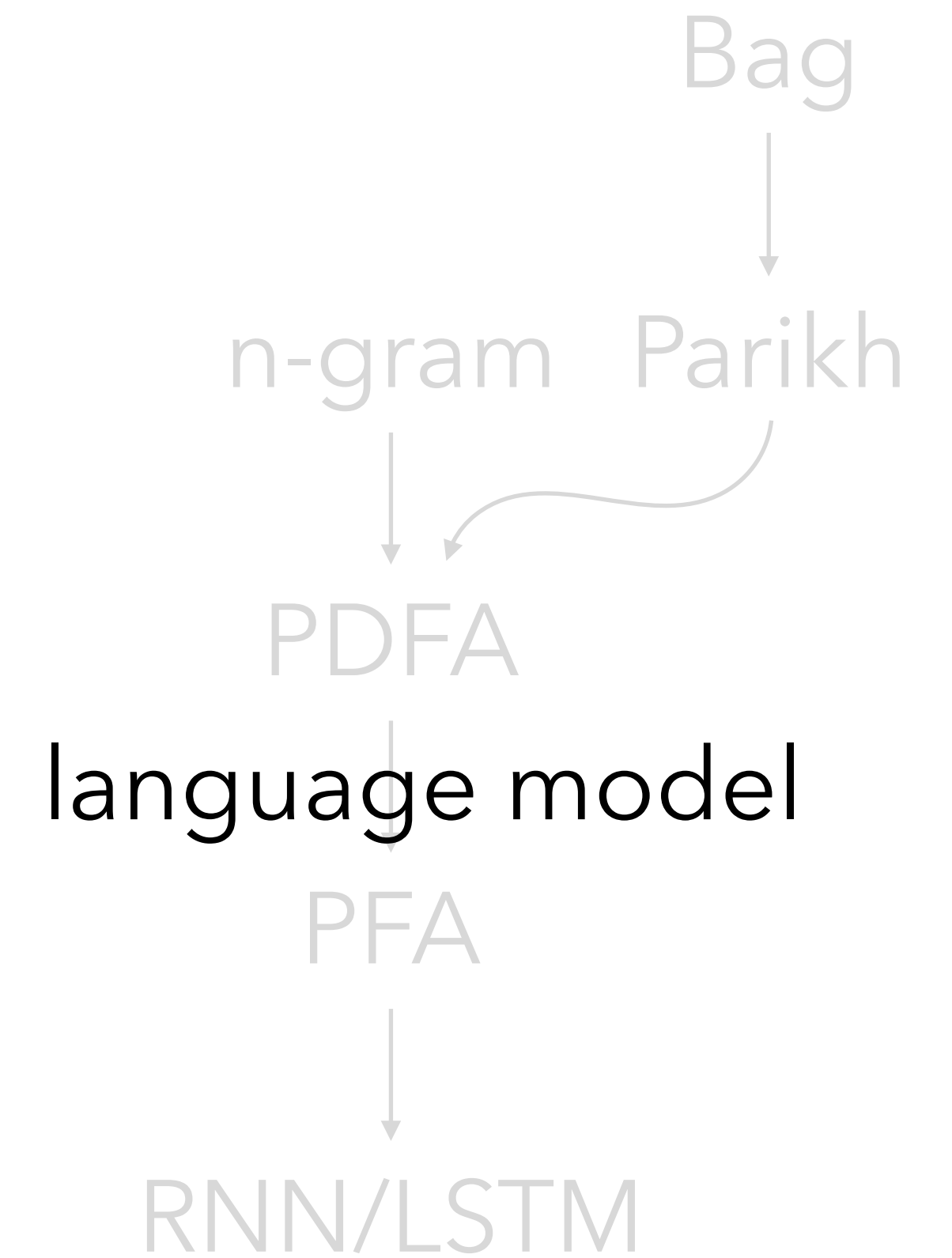
Language Models

Language models

Σ : finite alphabet of actions

$\Delta(S)$: distributions over S

$$p: \Sigma^* \rightarrow \Delta(\Sigma \cup \{\text{stop}\})$$



Language models

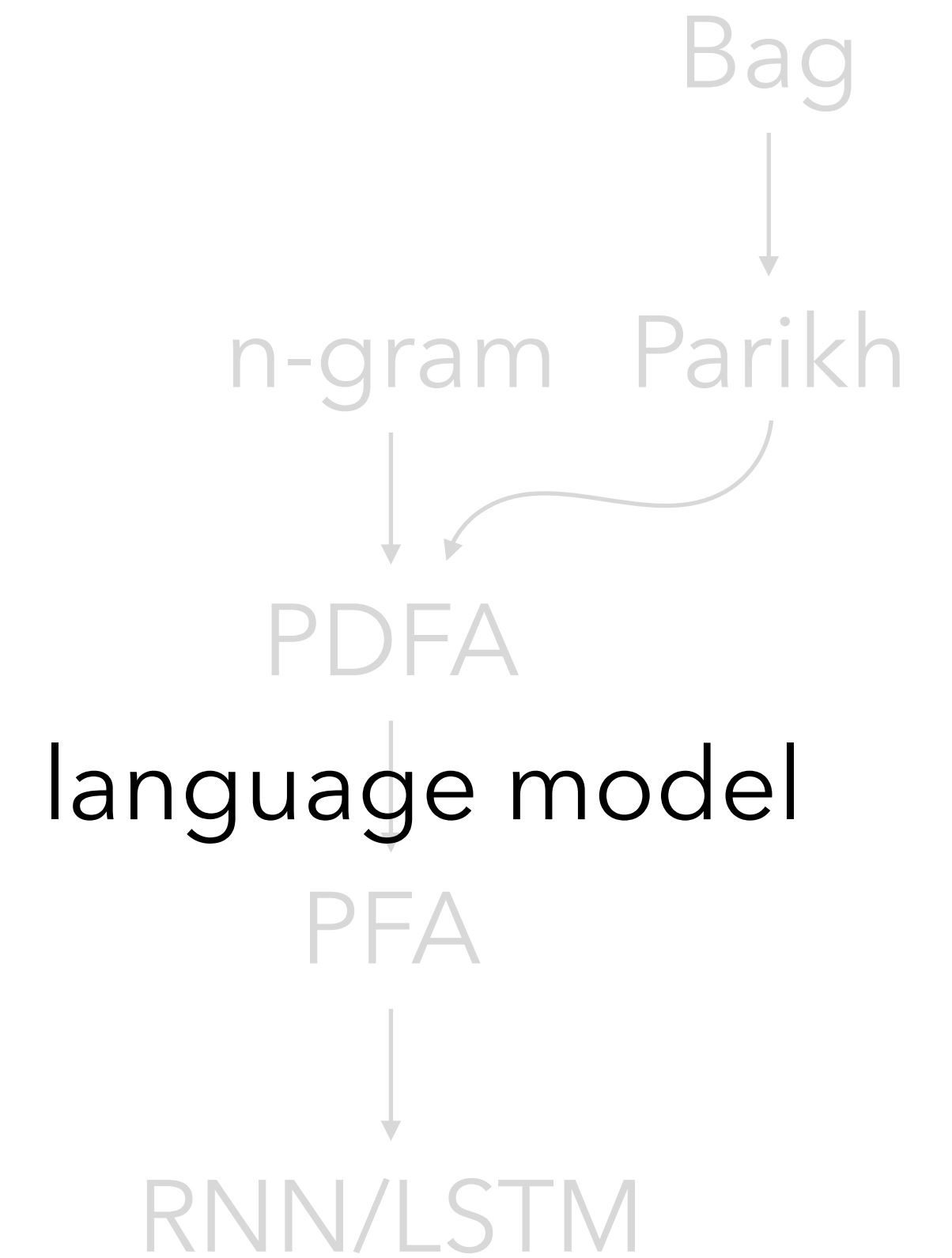
Σ : finite alphabet of actions

$\Delta(S)$: distributions over S

$w \in \Sigma^*$



$$p: \Sigma^* \rightarrow \Delta(\Sigma \cup \{\text{stop}\})$$



Language models

Σ : finite alphabet of actions

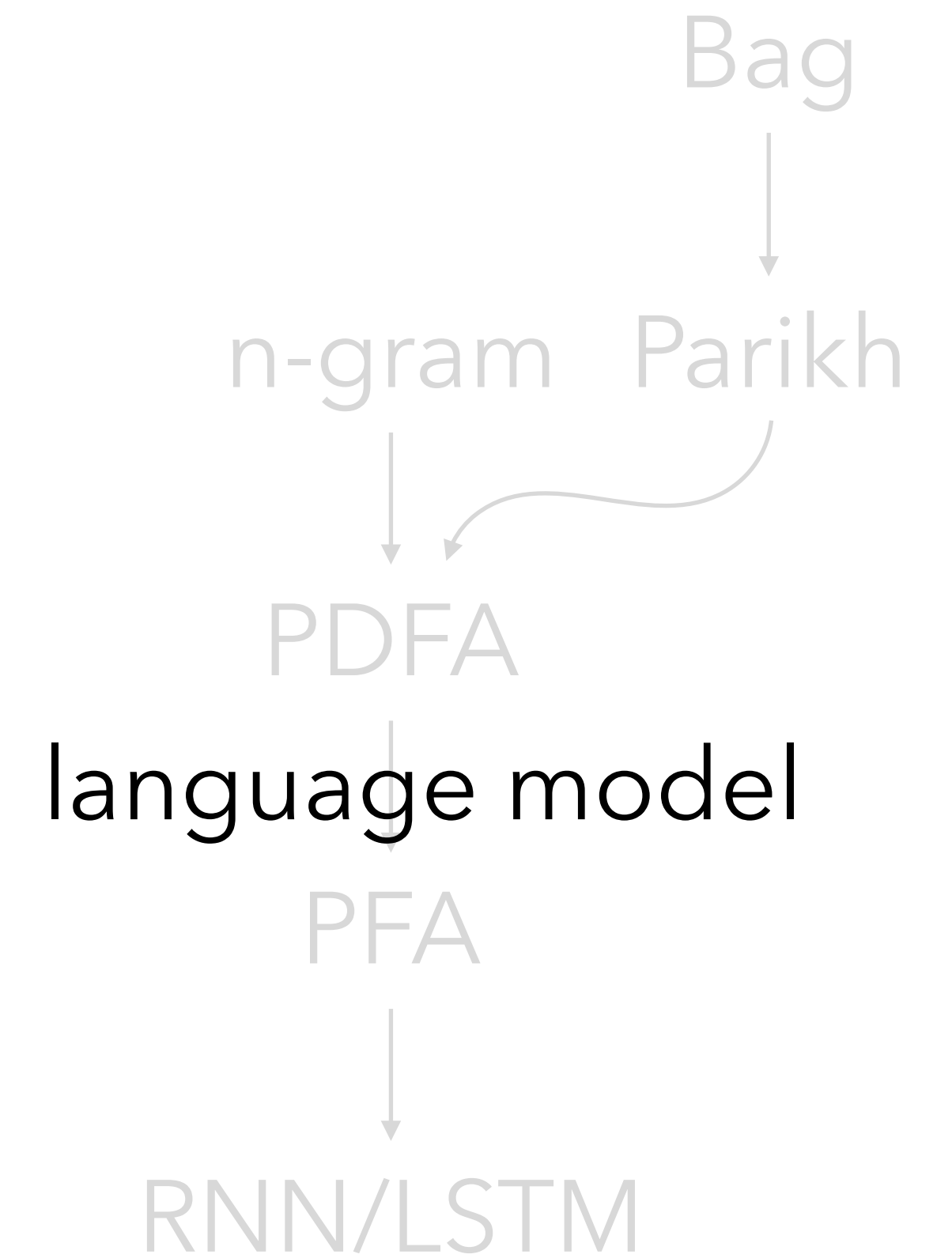
$\Delta(S)$: distributions over S

$$w \in \Sigma^*$$

inference

$$p: \Sigma^* \rightarrow \Delta(\Sigma \cup \{\text{stop}\})$$

$$p(\cdot | w) \in \Delta(\Sigma \cup \{\text{stop}\})$$



Language models

Σ : finite alphabet of actions

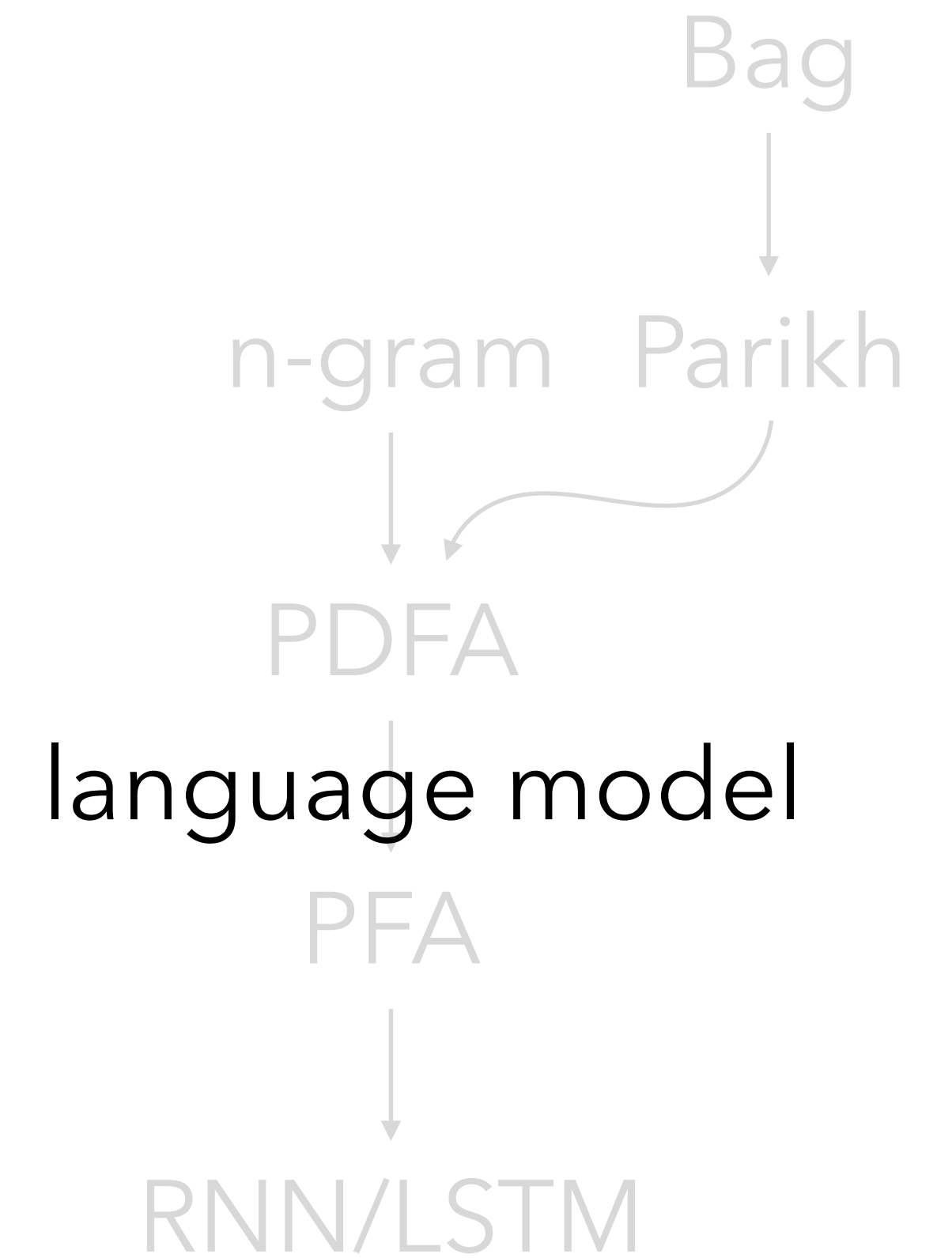
$\Delta(S)$: distributions over S

$$w \in \Sigma^*$$

inference

$$p: \Sigma^* \rightarrow \Delta(\Sigma \cup \{\text{stop}\})$$

$$p(\cdot | w) \in \Delta(\Sigma \cup \{\text{stop}\}) \longrightarrow \text{top } k\text{-actions/sample}$$



Language models

Σ : finite alphabet of actions

$\Delta(S)$: distributions over S

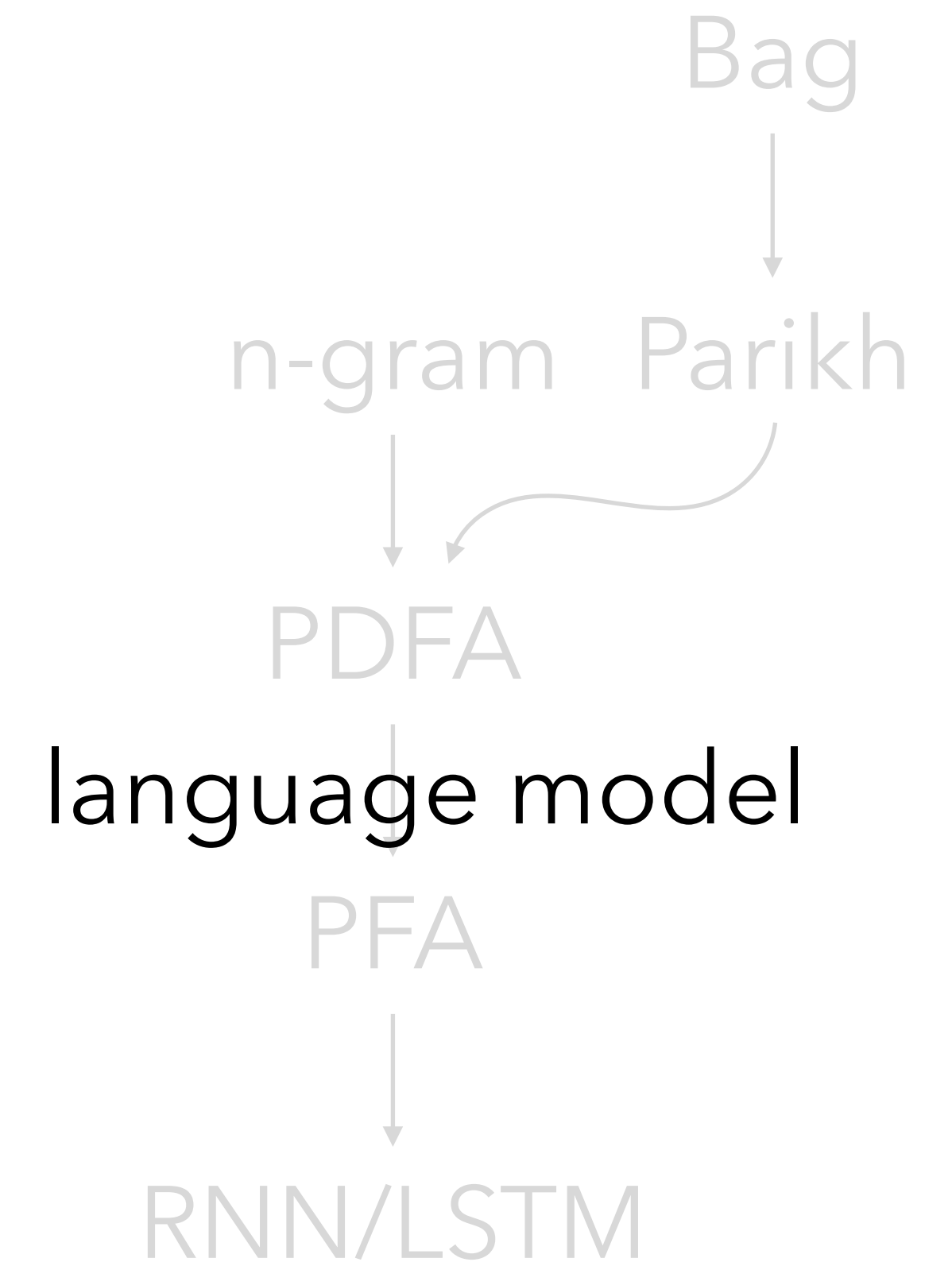
$$w \in \Sigma^*$$

inference

$$p: \Sigma^* \rightarrow \Delta(\Sigma \cup \{\text{stop}\})$$

autoregression (under certain conditions)

$$\text{Pr} \in \Delta(\Sigma^*)$$



$$p(\cdot | w) \in \Delta(\Sigma \cup \{\text{stop}\}) \longrightarrow \text{top } k\text{-actions/sample}$$

Language models

Σ : finite alphabet of actions

$\Delta(S)$: distributions over S

$$w \in \Sigma^*$$

inference

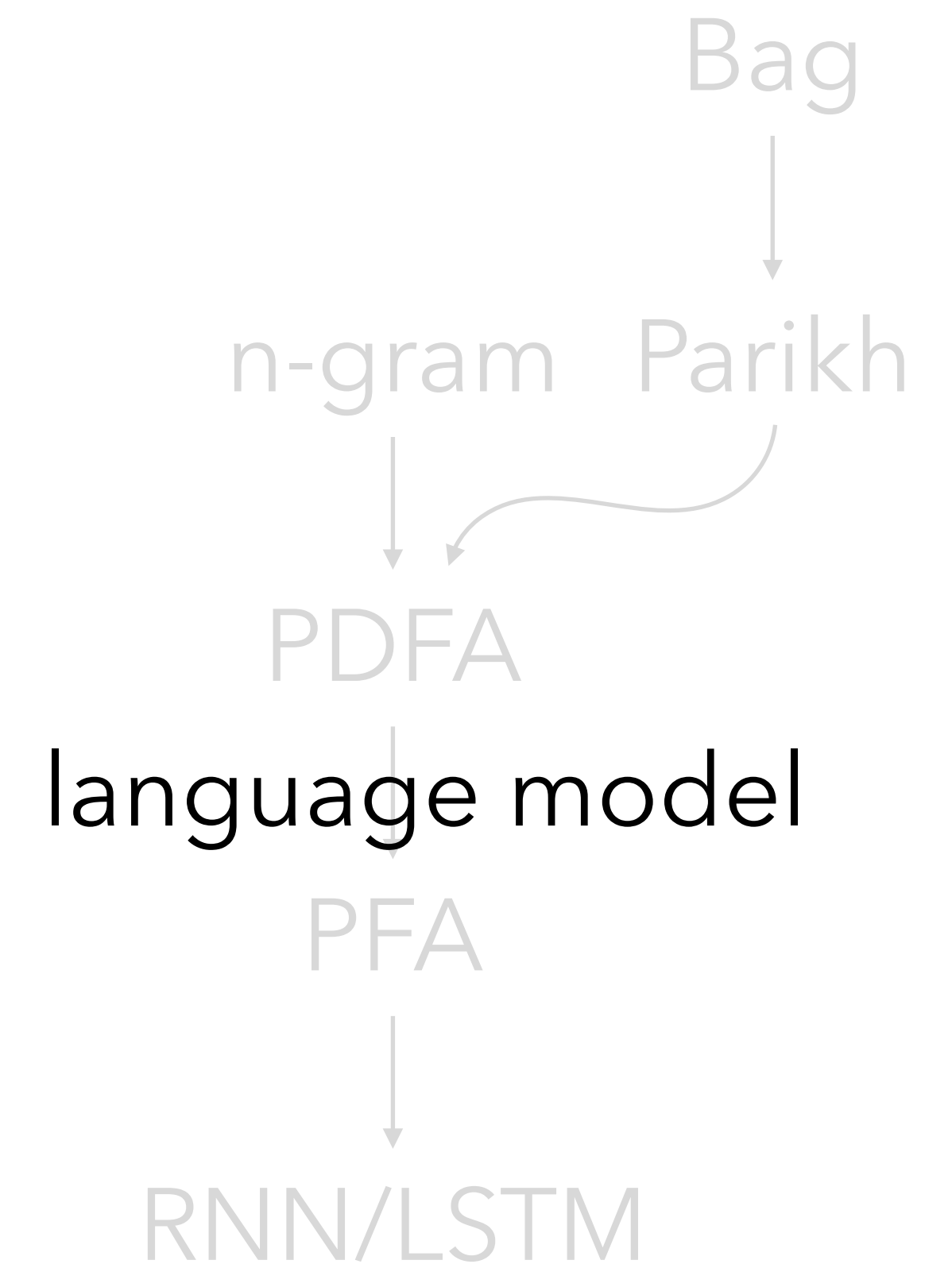
$$p: \Sigma^* \rightarrow \Delta(\Sigma \cup \{\text{stop}\})$$

autoregression (under certain conditions)

$$\text{Pr} \in \Delta(\Sigma^*)$$

$$\text{Pr}(a_1 \dots a_n) = \prod_{1 \leq i \leq n} p(a_i | a_1 \dots a_{i-1}) \cdot p(\text{stop} | a_1 \dots a_n)$$

$$p(\cdot | w) \in \Delta(\Sigma \cup \{\text{stop}\}) \longrightarrow \text{top } k\text{-actions/sample}$$



Language models

Σ : finite alphabet of actions

$\Delta(S)$: distributions over S

$$w \in \Sigma^*$$

inference

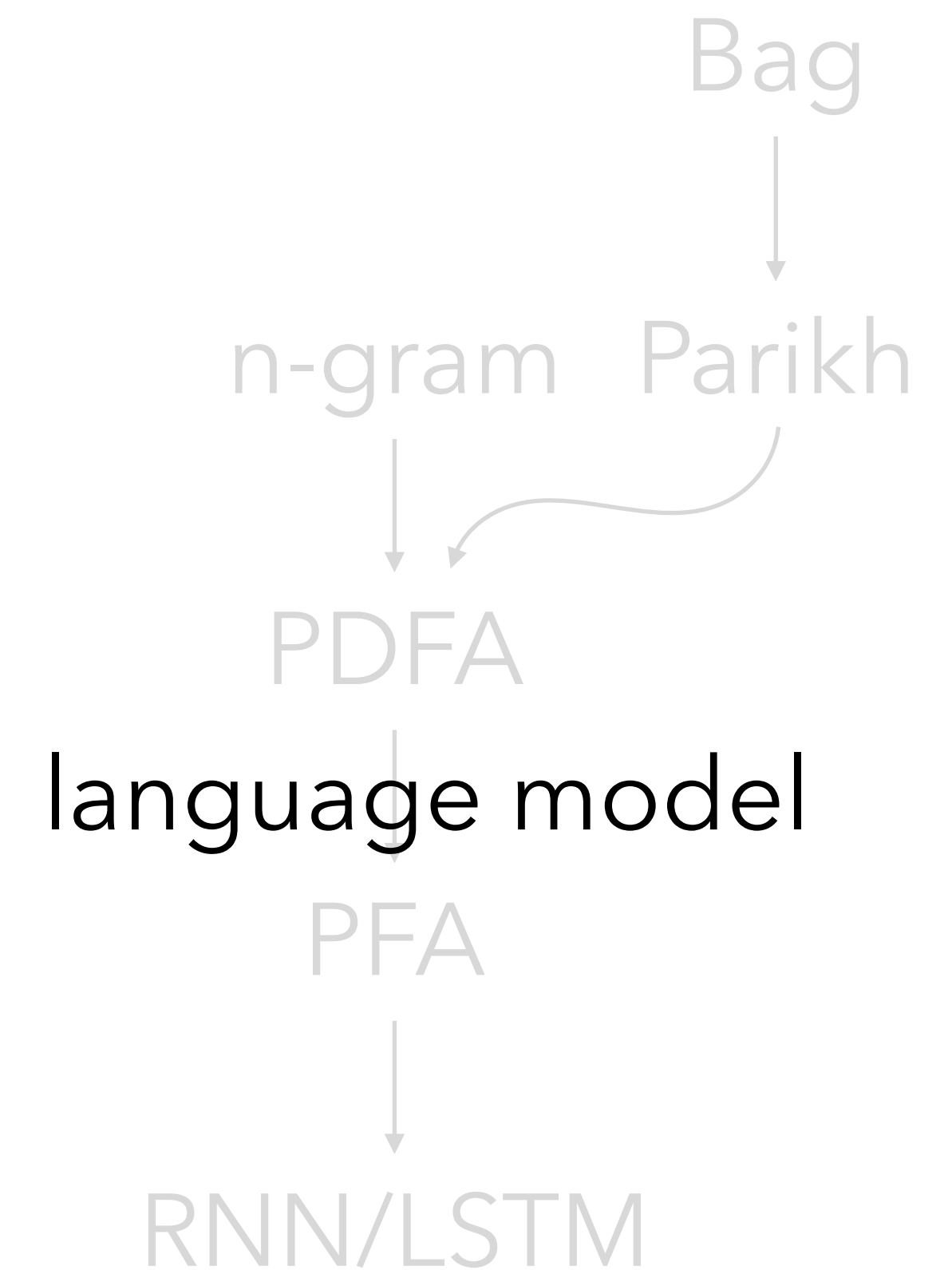
$$p: \Sigma^* \rightarrow \Delta(\Sigma \cup \{\text{stop}\})$$

autoregression (under certain conditions)

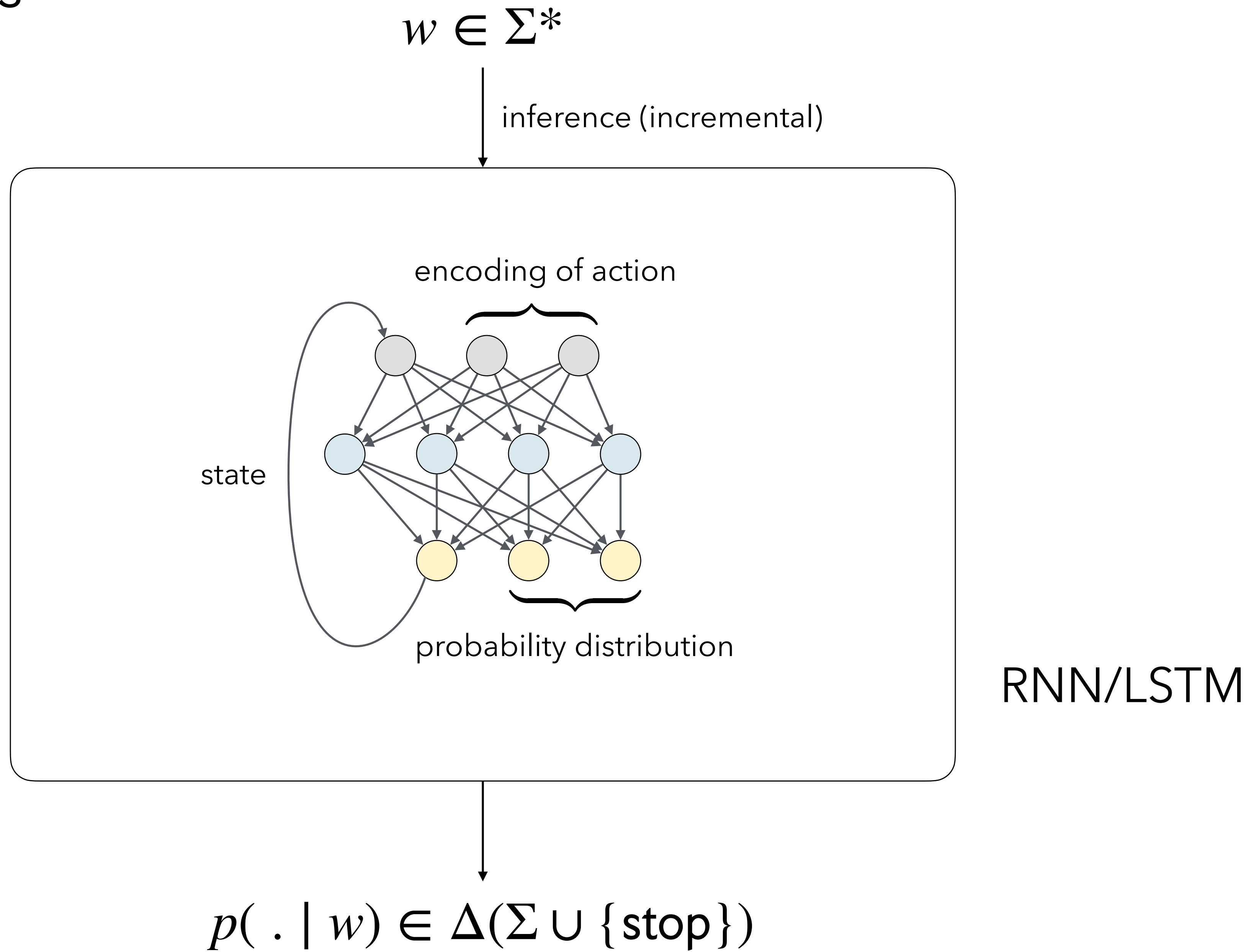
$$\text{Pr} \in \Delta(\Sigma^*)$$

$$\text{Pr}(a_1 \dots a_n) = \prod_{1 \leq i \leq n} p(a_i | a_1 \dots a_{i-1}) \cdot p(\text{stop} | a_1 \dots a_n)$$

$$p(\cdot | w) \in \Delta(\Sigma \cup \{\text{stop}\}) \longrightarrow \text{top } k\text{-actions/sample}$$

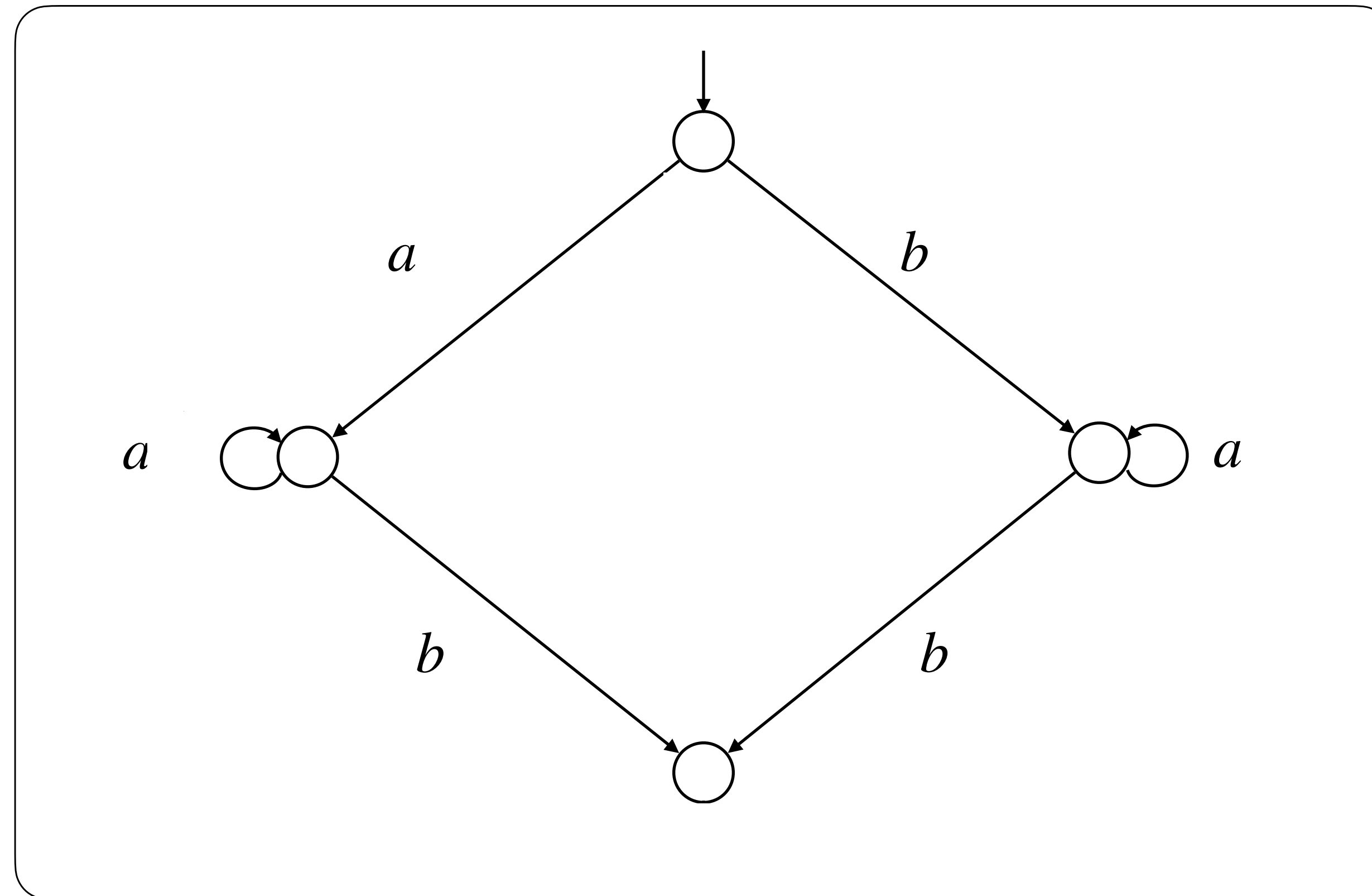


Language models



Language models

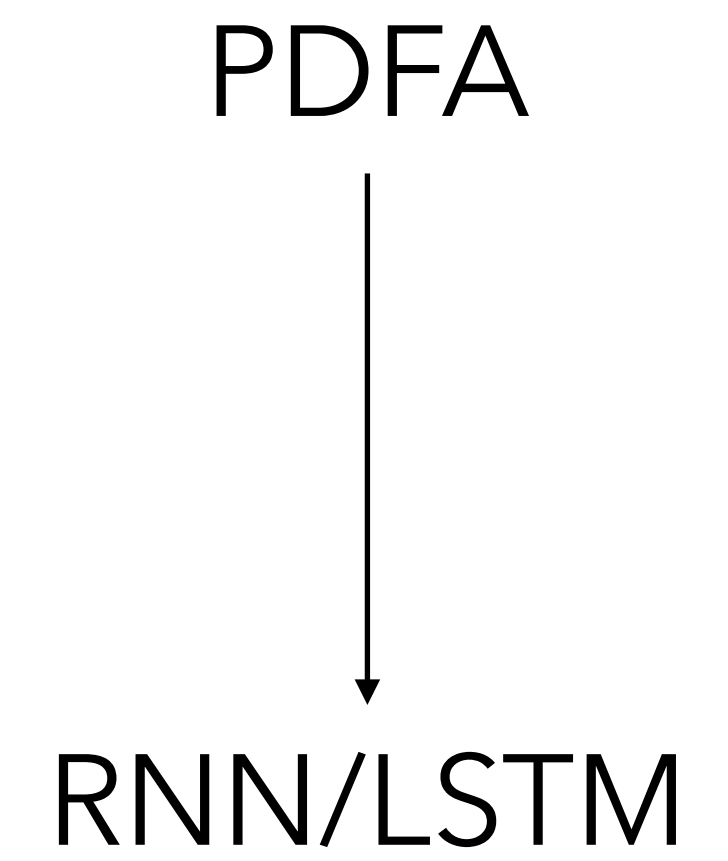
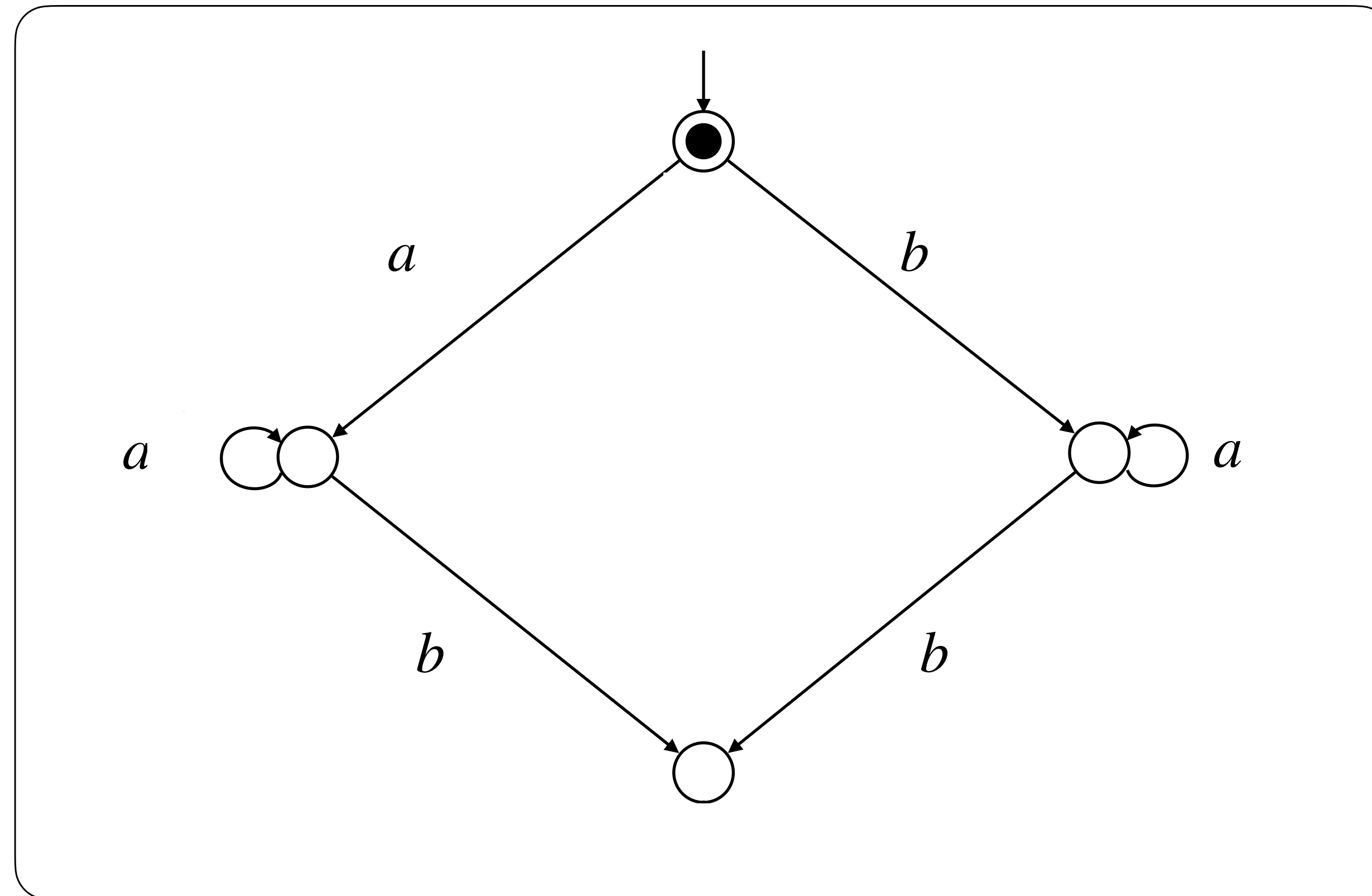
$\Sigma = \{a, b\}$



PDFA
↓
RNN/LSTM

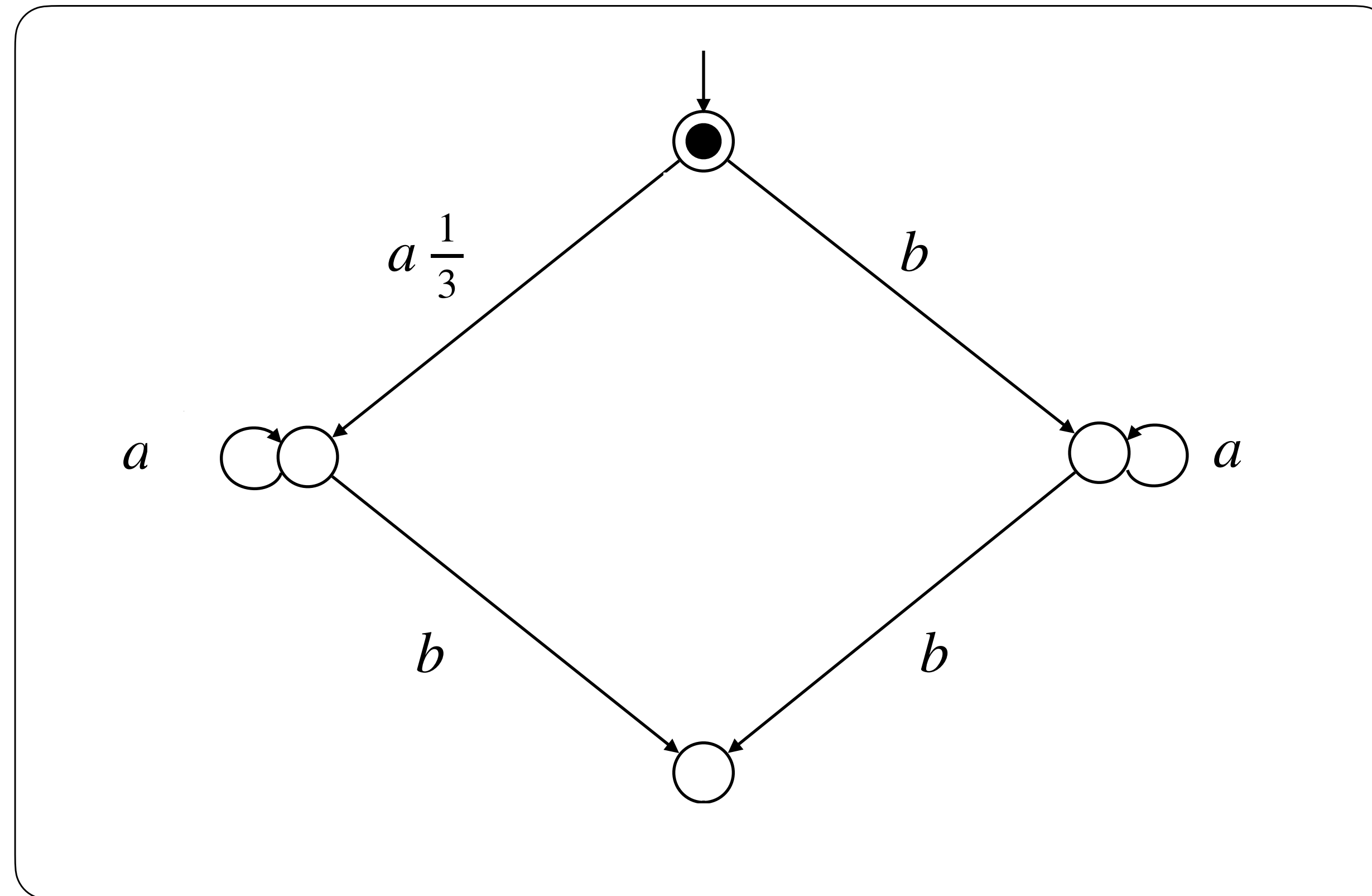
Language models

$\Sigma = \{a, b\}$



Language models

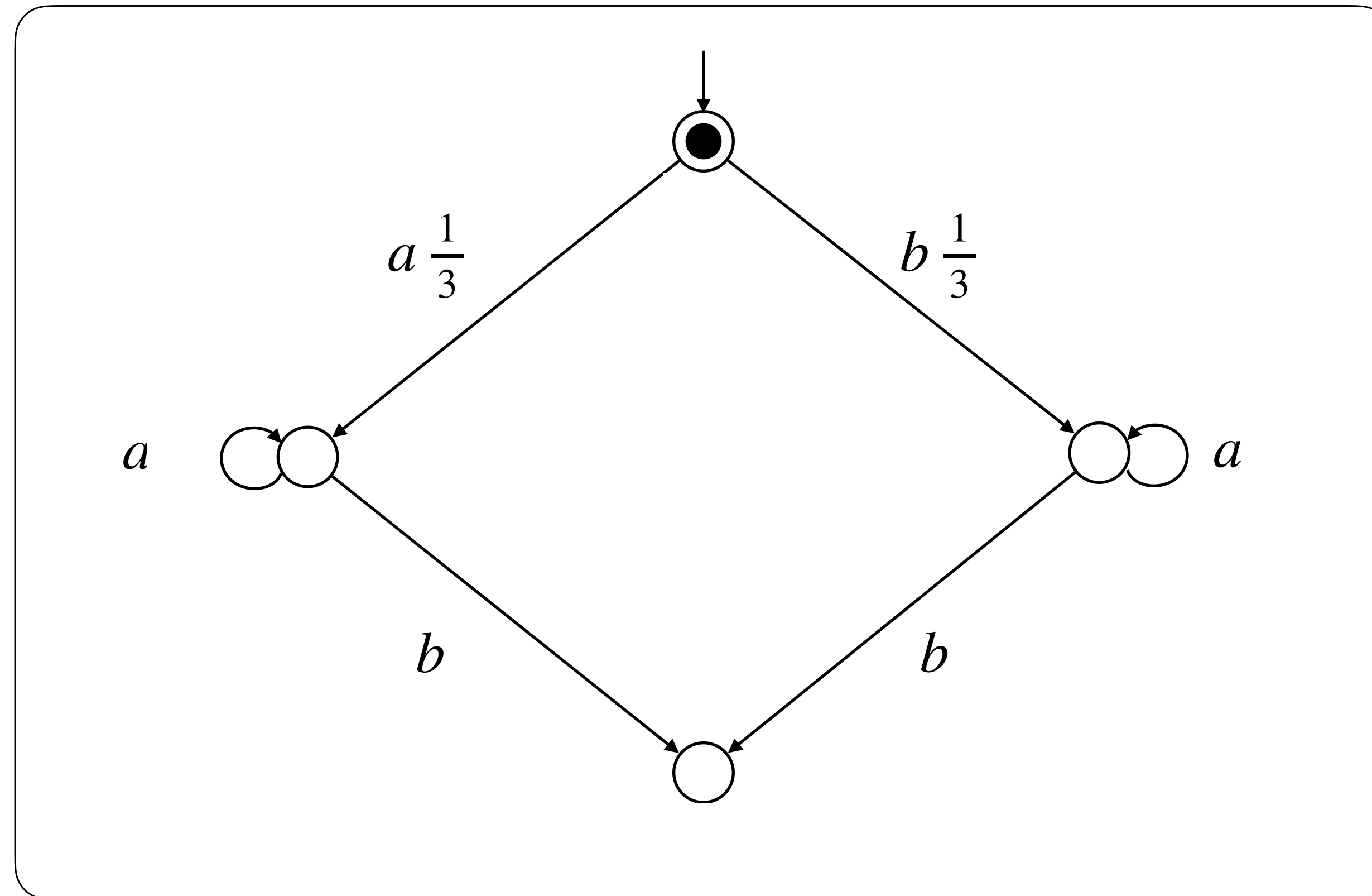
$\Sigma = \{a, b\}$



PDFA
↓
RNN/LSTM

Language models

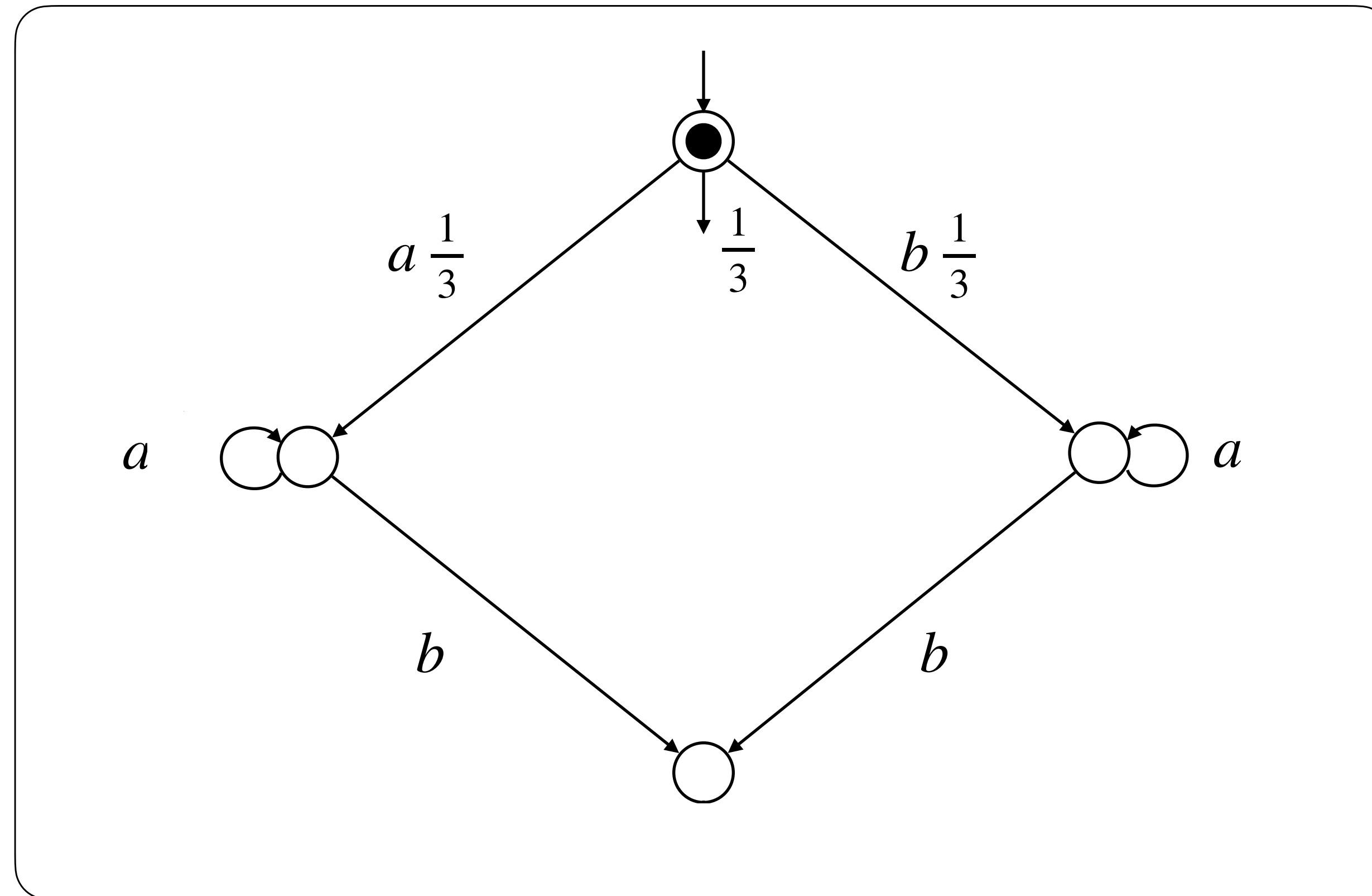
$\Sigma = \{a, b\}$



PDFA
↓
RNN/LSTM

Language models

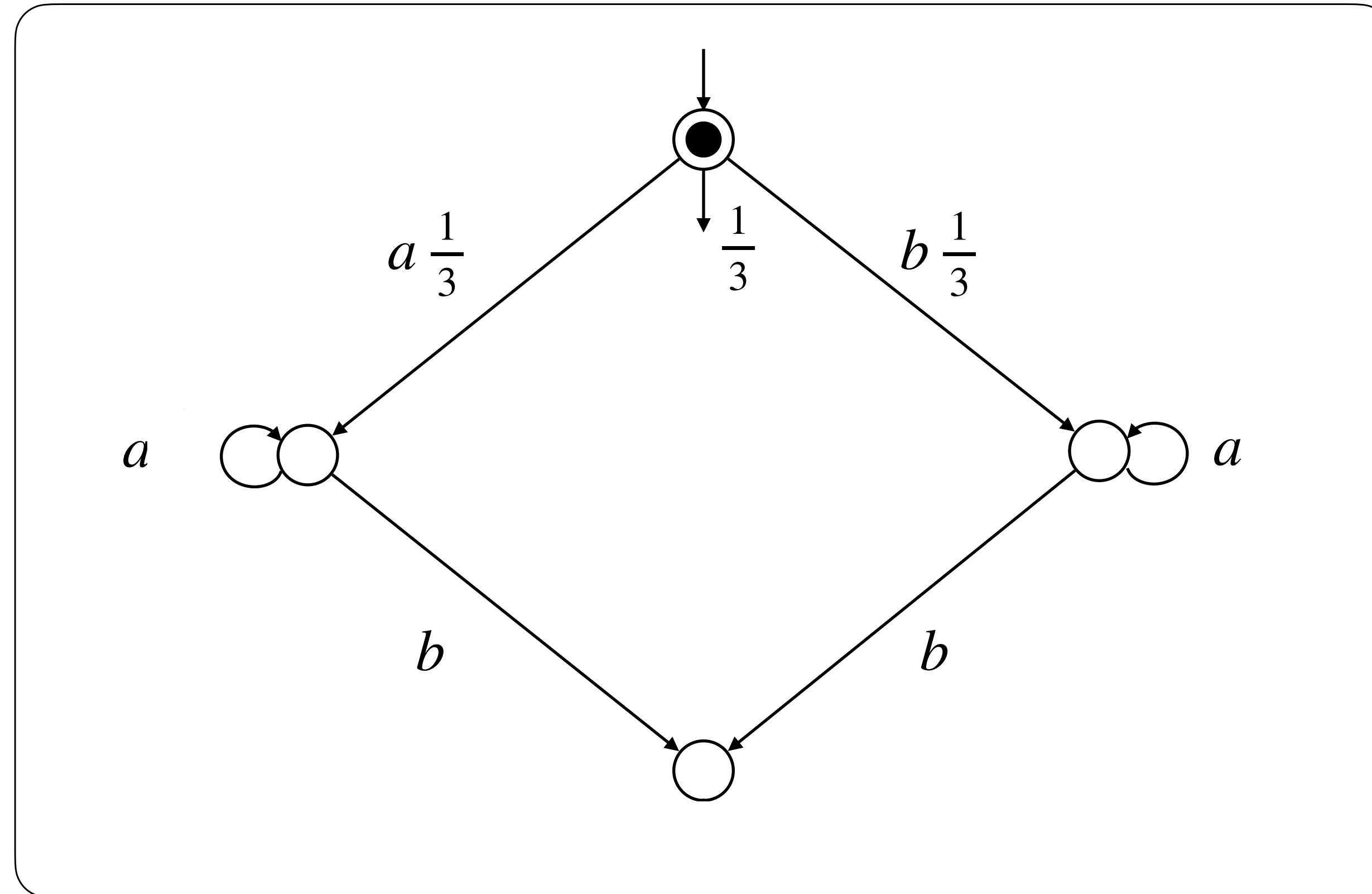
$\Sigma = \{a, b\}$



PDFFA
↓
RNN/LSTM

Language models

$\Sigma = \{a, b\}$

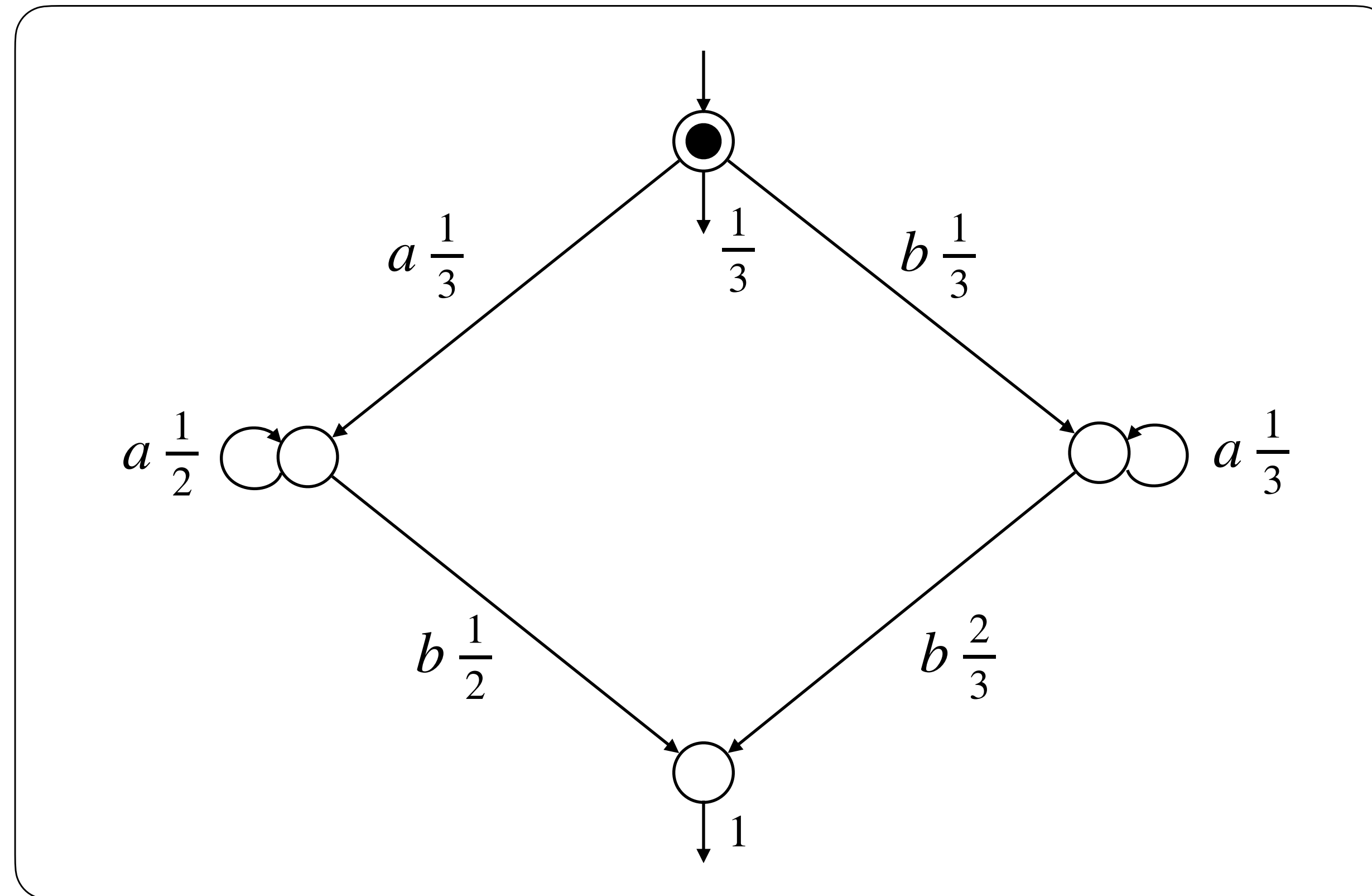


PDFA
↓
RNN/LSTM

$$p(a \mid \varepsilon) = p(b \mid \varepsilon) = p(\text{stop} \mid \varepsilon) = \frac{1}{3}$$

Language models

$\Sigma = \{a, b\}$



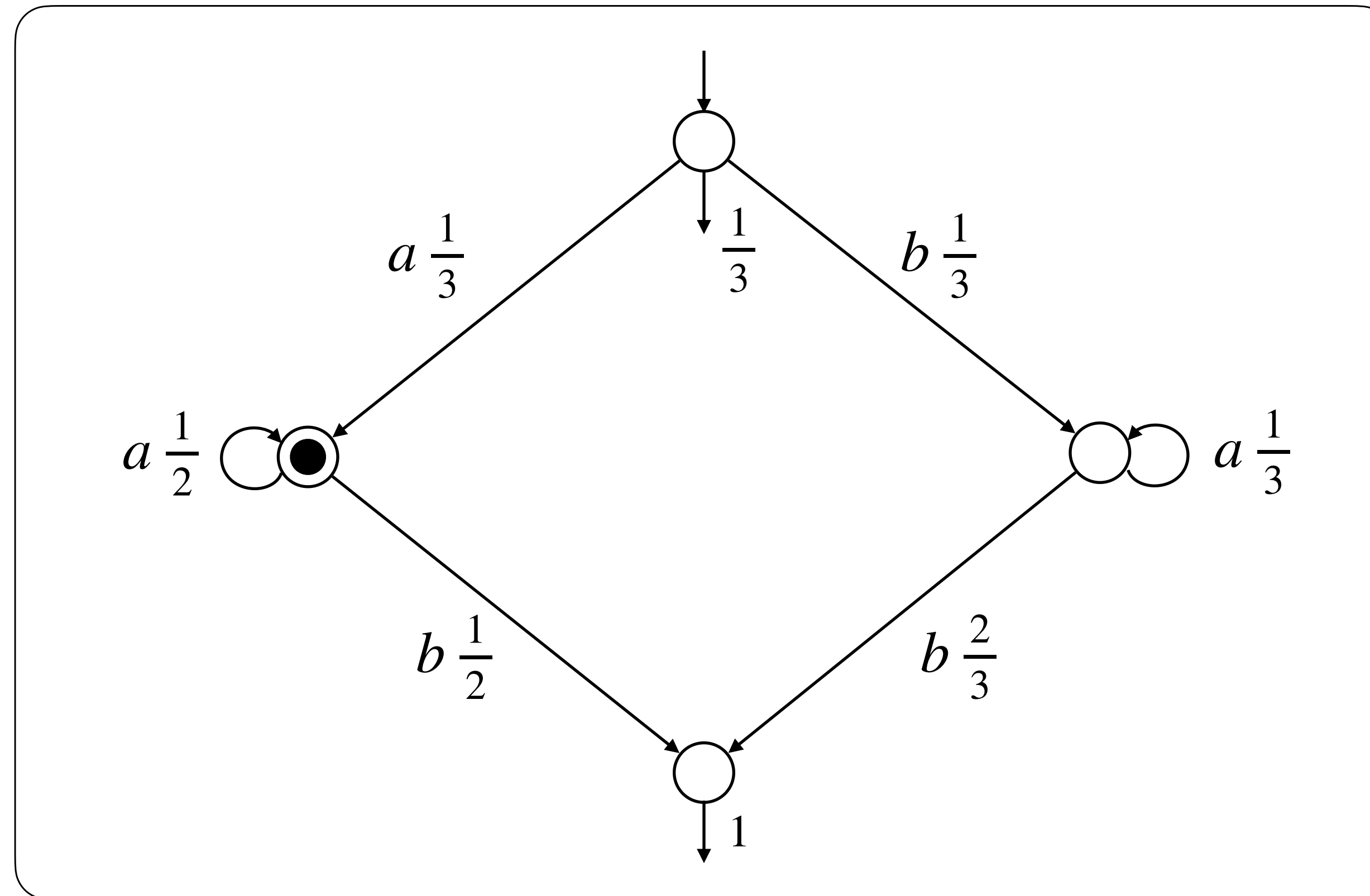
PDFFA
↓
RNN/LSTM

$$p(a \mid \varepsilon) = p(b \mid \varepsilon) = p(\text{stop} \mid \varepsilon) = \frac{1}{3}$$

$$p(a \mid aa) = p(b \mid aa) = \frac{1}{2}$$

Language models

$\Sigma = \{a, b\}$



PDFFA
↓
RNN/LSTM

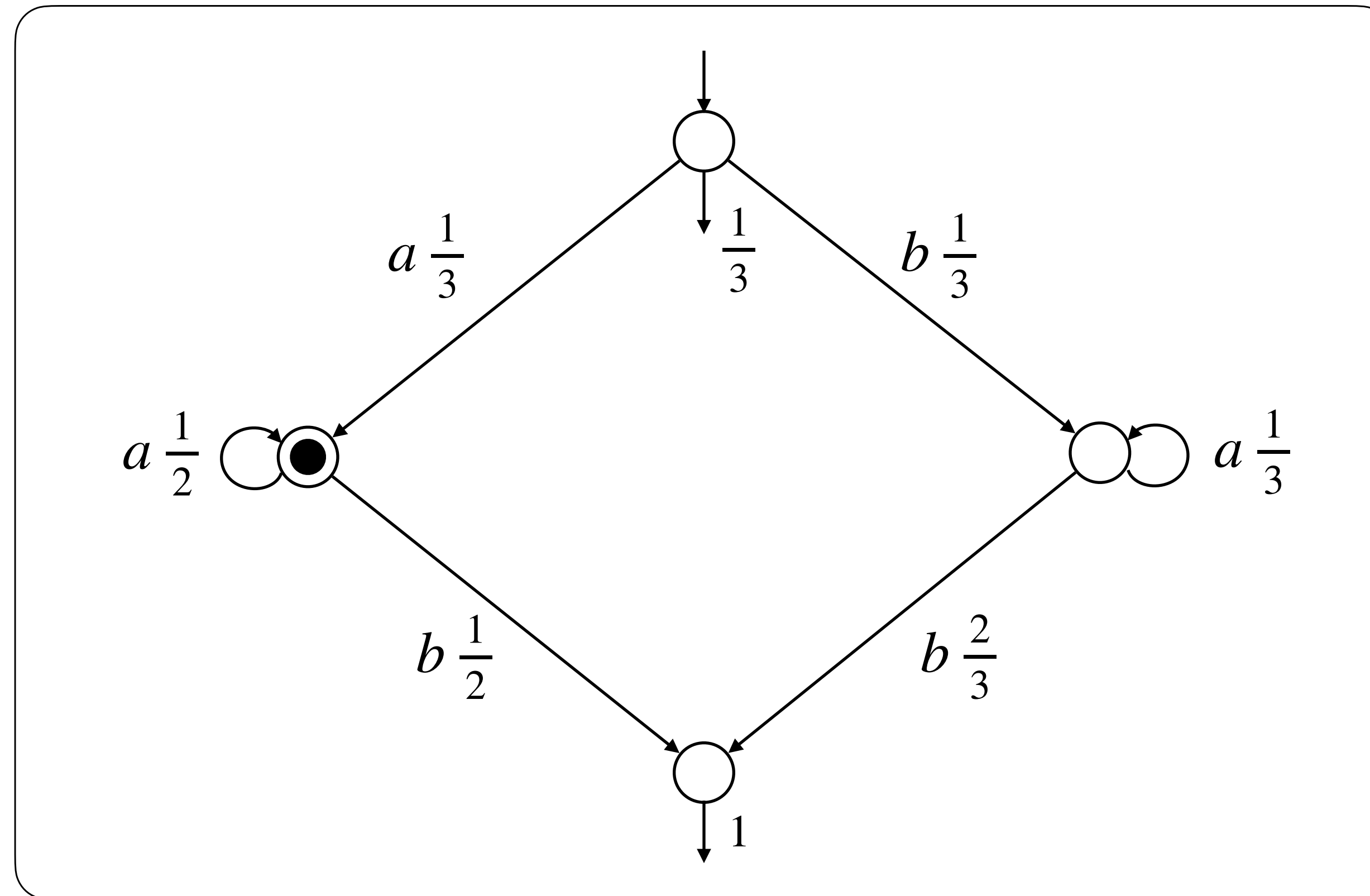
$$p(a \mid \varepsilon) = p(b \mid \varepsilon) = p(\text{stop} \mid \varepsilon) = \frac{1}{3}$$

$$p(a \mid aa) = p(b \mid aa) = \frac{1}{2}$$

Language models

$$\Sigma = \{a, b\}$$

incremental:
when a new action comes in,
we only update current state



PDFFA
↓
RNN/LSTM

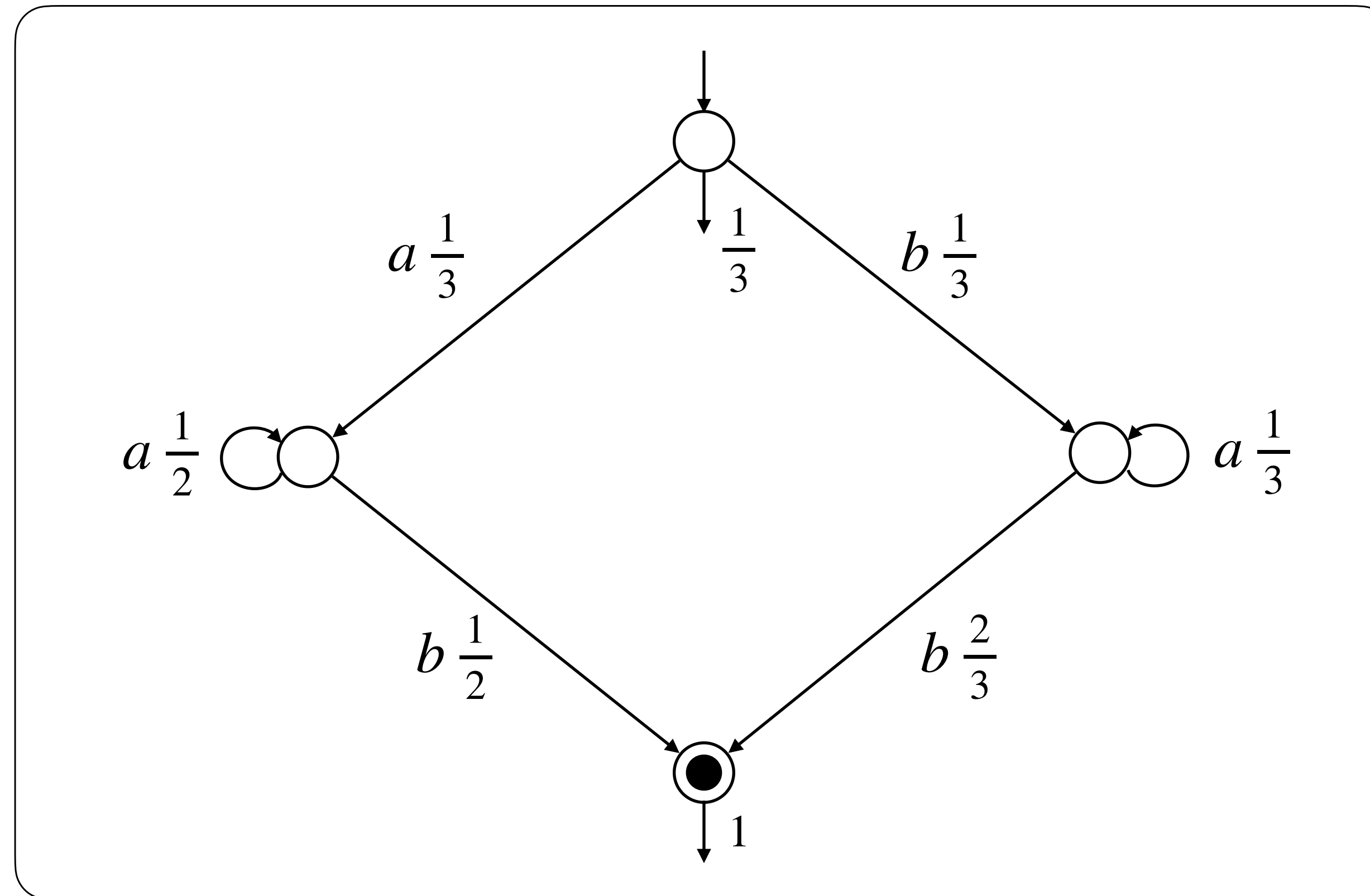
$$p(a \mid \varepsilon) = p(b \mid \varepsilon) = p(\text{stop} \mid \varepsilon) = \frac{1}{3}$$

$$p(a \mid aa) = p(b \mid aa) = \frac{1}{2}$$

Language models

$$\Sigma = \{a, b\}$$

incremental:
when a new action comes in,
we only update current state



PDFFA
↓
RNN/LSTM

$$p(a \mid \varepsilon) = p(b \mid \varepsilon) = p(\text{stop} \mid \varepsilon) = \frac{1}{3}$$

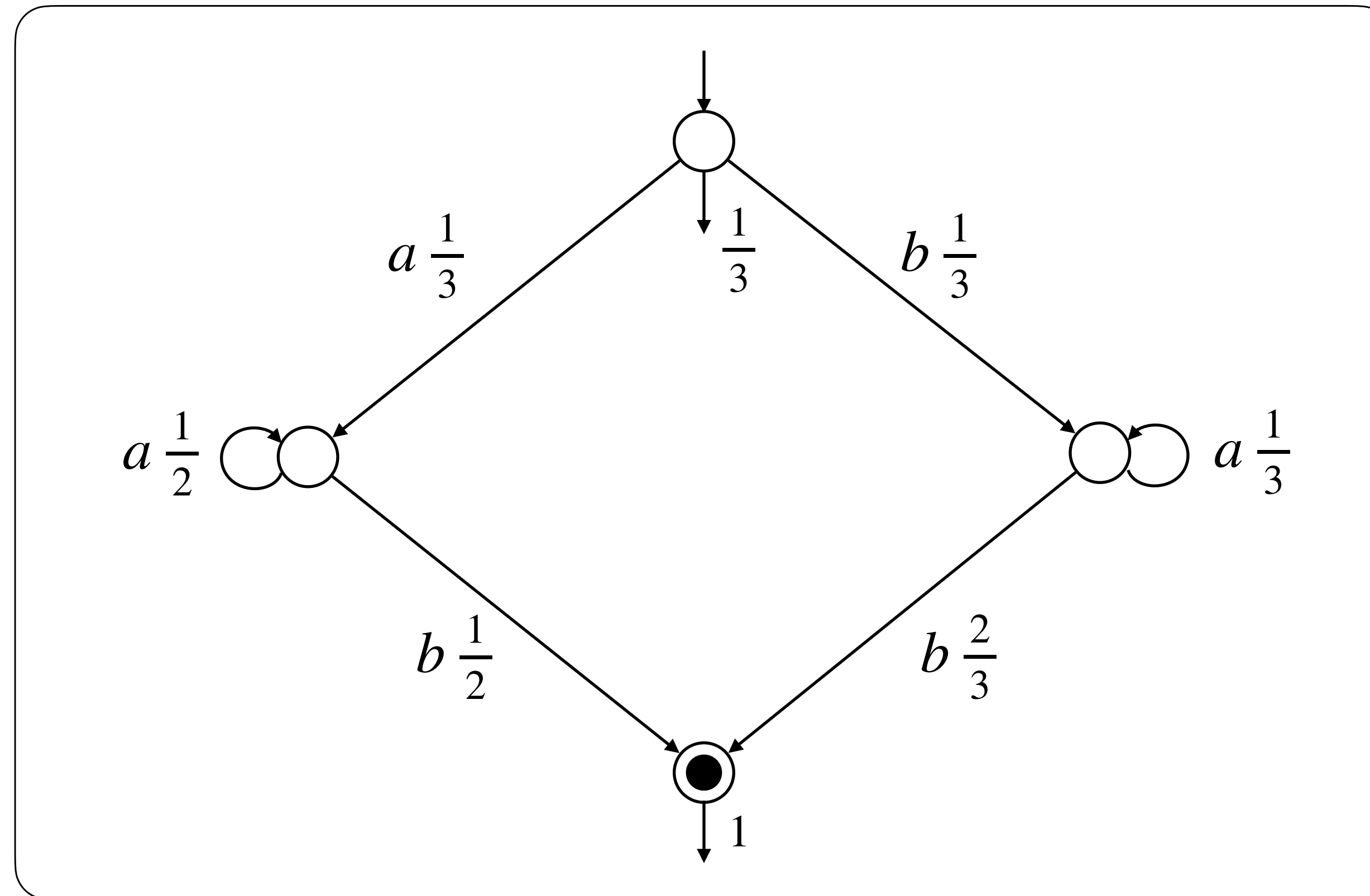
$$p(a \mid aa) = p(b \mid aa) = \frac{1}{2}$$

Language models

$$\Sigma = \{a, b\}$$

incremental:
when a new action comes in,
we only update current state

$$p(\text{stop} \mid aab) = 1$$



PDFA
↓
RNN/LSTM

$$p(a \mid \varepsilon) = p(b \mid \varepsilon) = p(\text{stop} \mid \varepsilon) = \frac{1}{3}$$

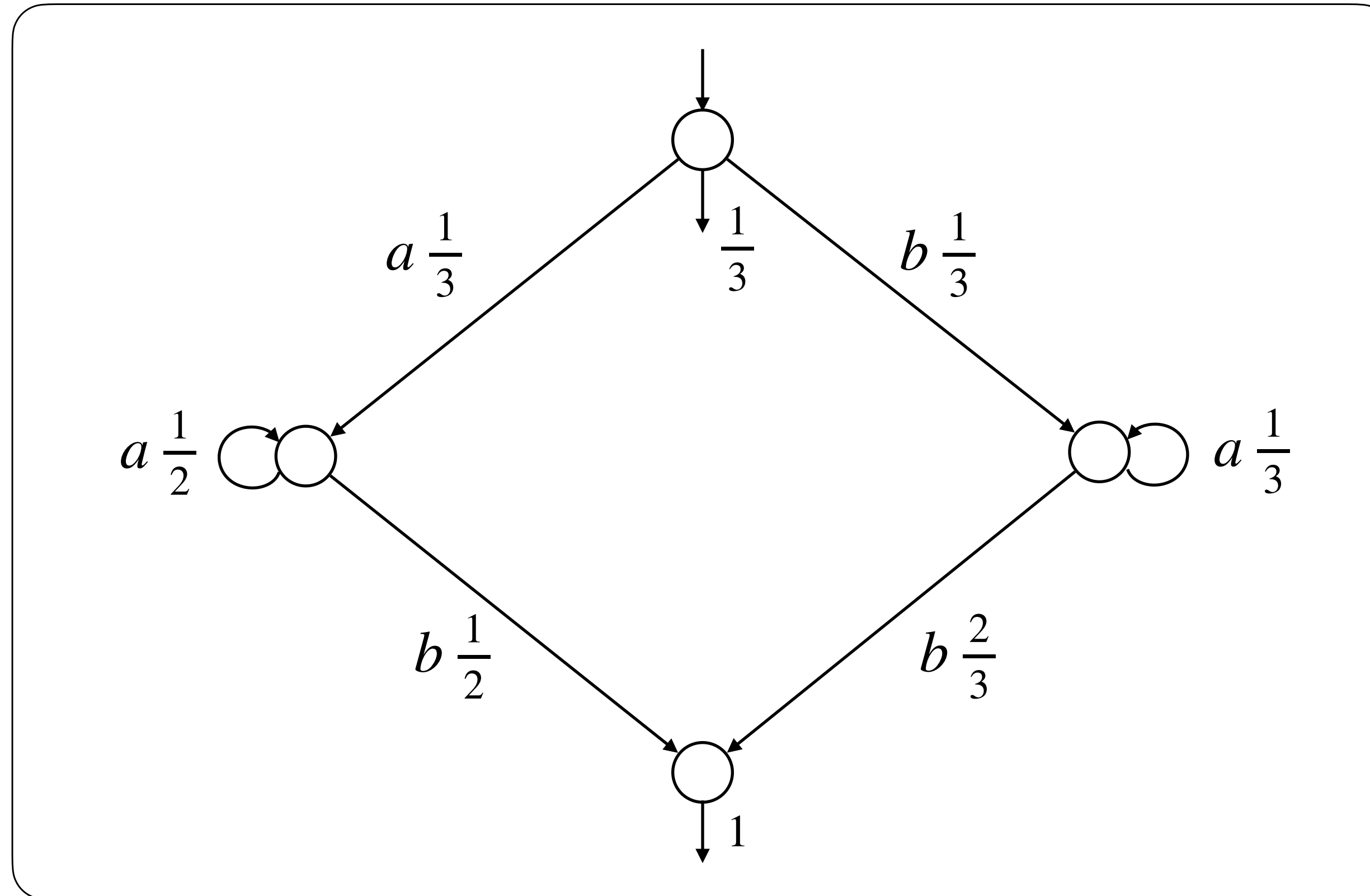
$$p(a \mid aa) = p(b \mid aa) = \frac{1}{2}$$

Language models

$$\Sigma = \{a, b\}$$

incremental:
when a new action comes in,
we only update current state

$$p(\text{stop} \mid aab) = 1$$



PDFFA
↓
RNN/LSTM

$$p(a \mid \varepsilon) = p(b \mid \varepsilon) = p(\text{stop} \mid \varepsilon) = \frac{1}{3}$$

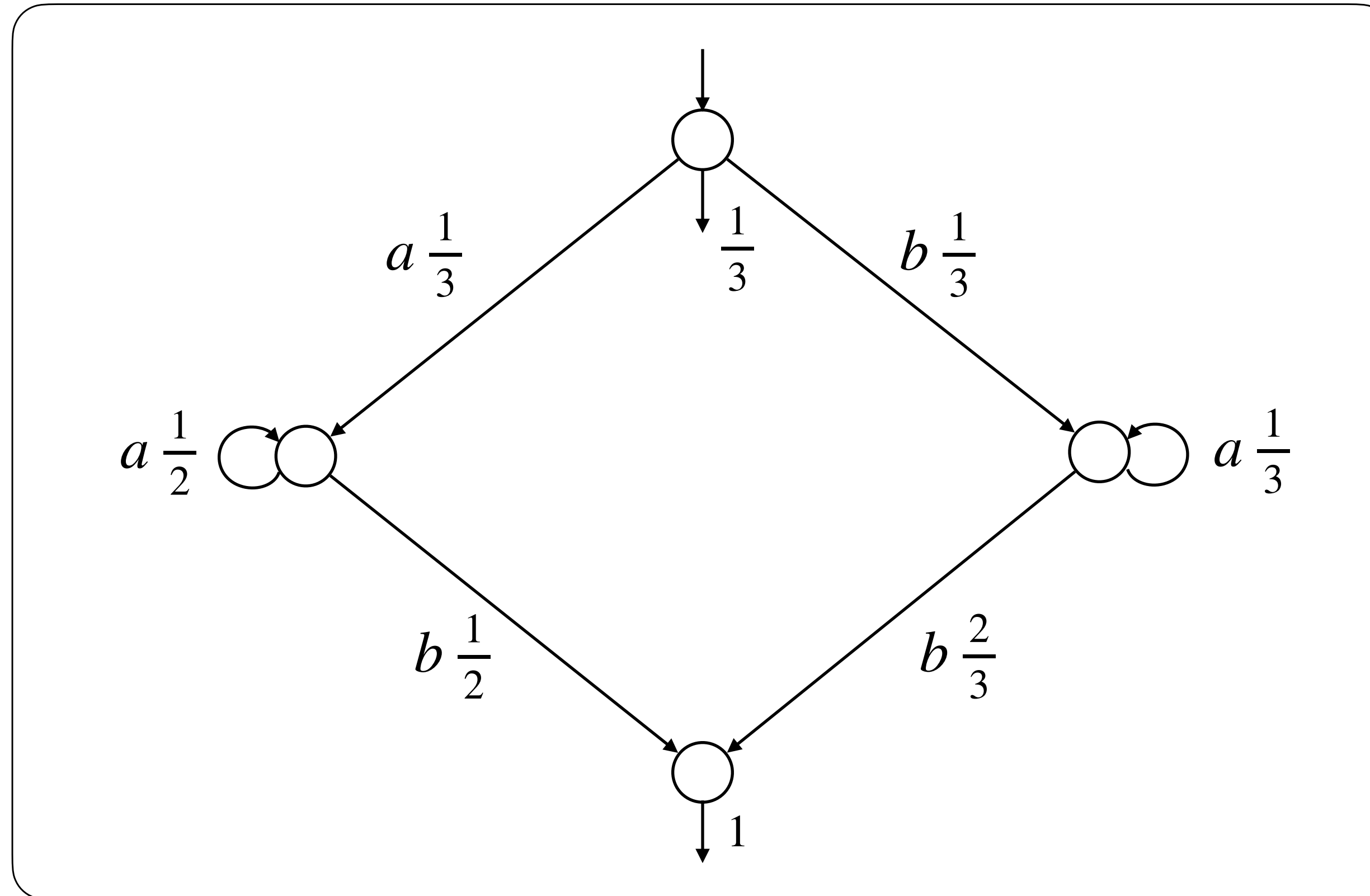
$$p(a \mid aa) = p(b \mid aa) = \frac{1}{2}$$

Language models

$$\Sigma = \{a, b\}$$

incremental:
when a new action comes in,
we only update current state

$$p(\text{stop} \mid aab) = 1$$



probabilities depend
on last $n - 1$ actions

n -gram



PDFFA



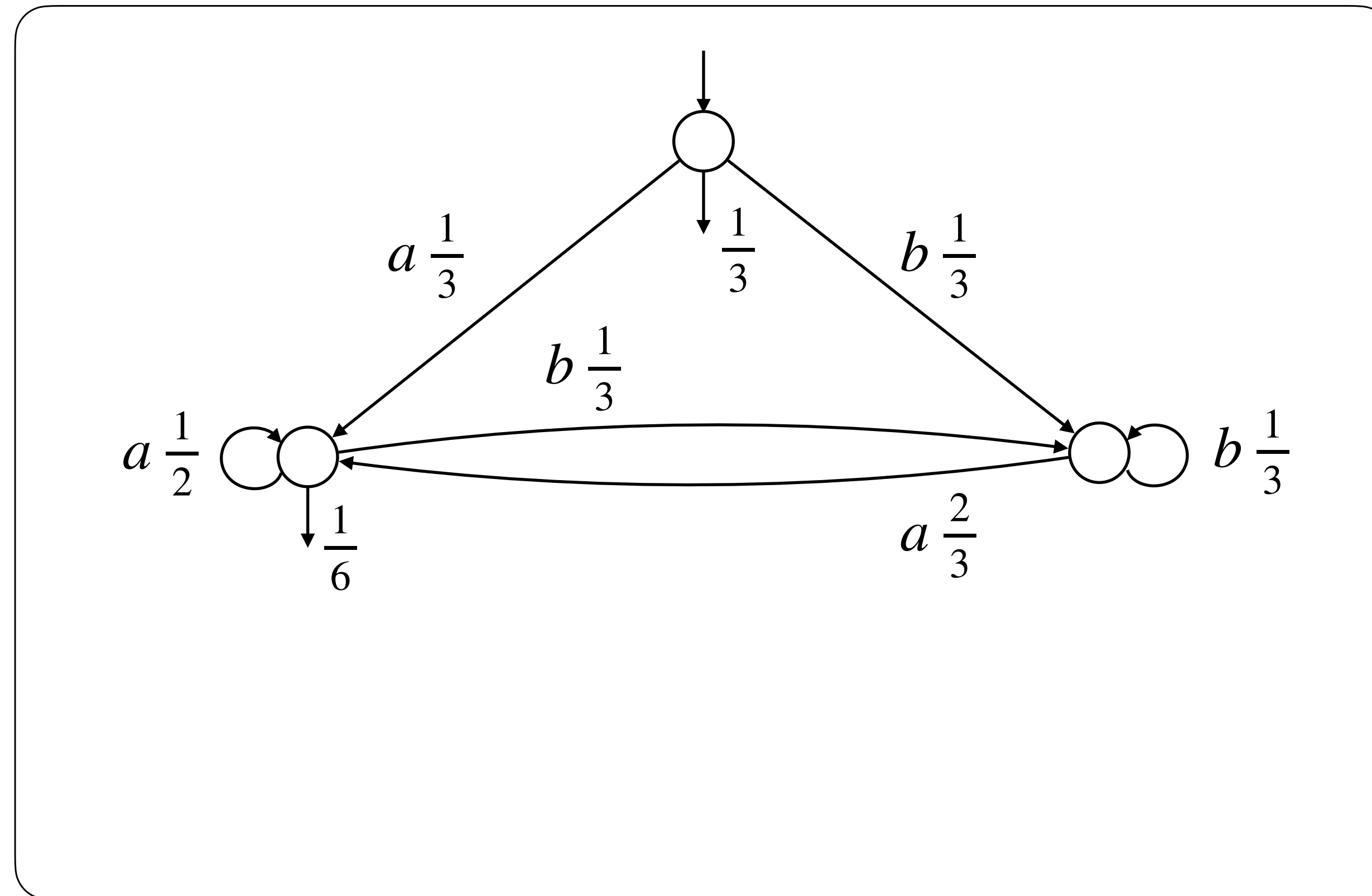
RNN/LSTM

$$p(a \mid \varepsilon) = p(b \mid \varepsilon) = p(\text{stop} \mid \varepsilon) = \frac{1}{3}$$

$$p(a \mid aa) = p(b \mid aa) = \frac{1}{2}$$

Language models

$$\Sigma = \{a, b\}$$



probabilities depend
on last action

2-gram



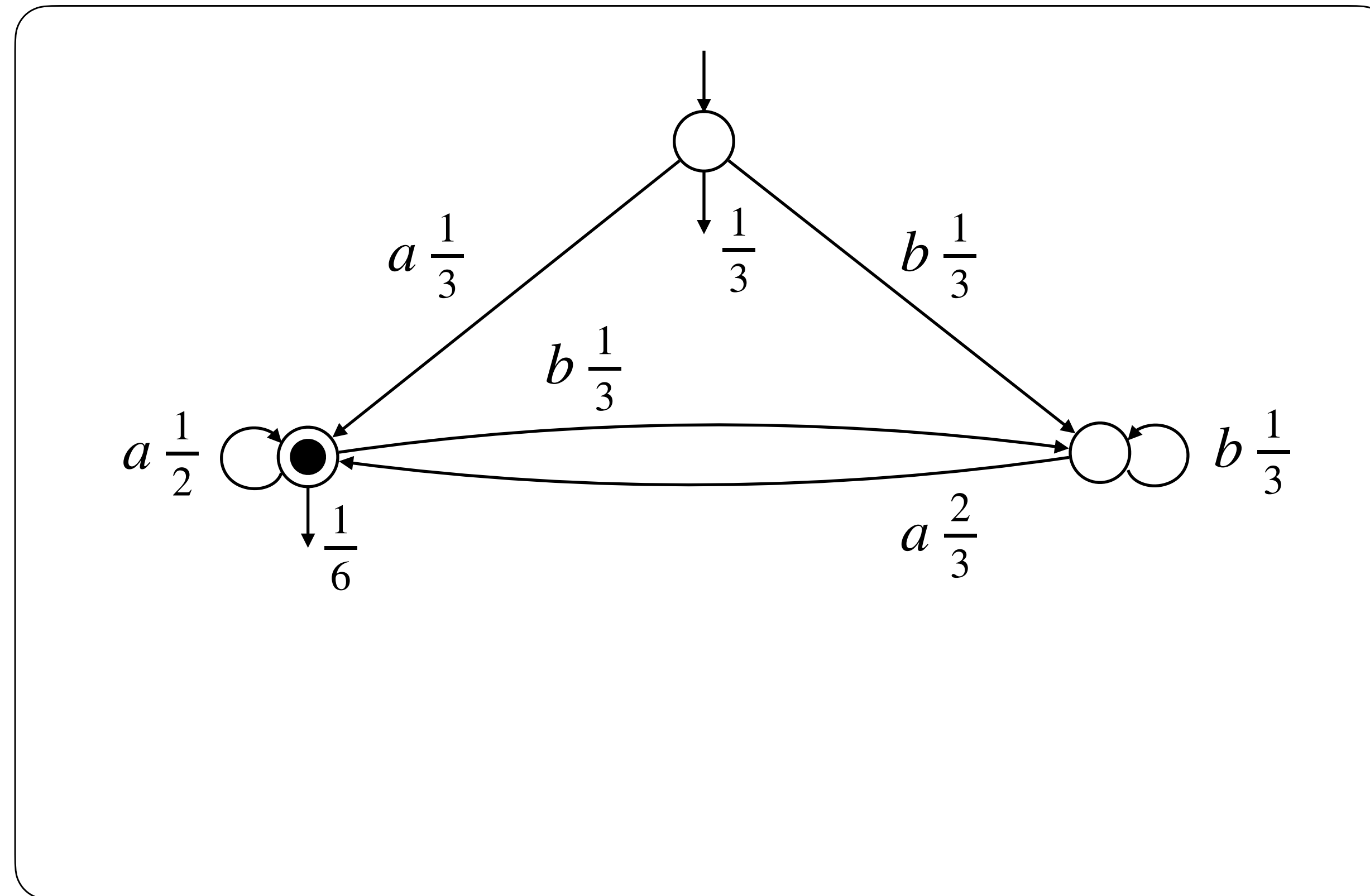
PDFFA



RNN/LSTM

Language models

$$\Sigma = \{a, b\}$$



probabilities depend
on last action

2-gram



PDFFA

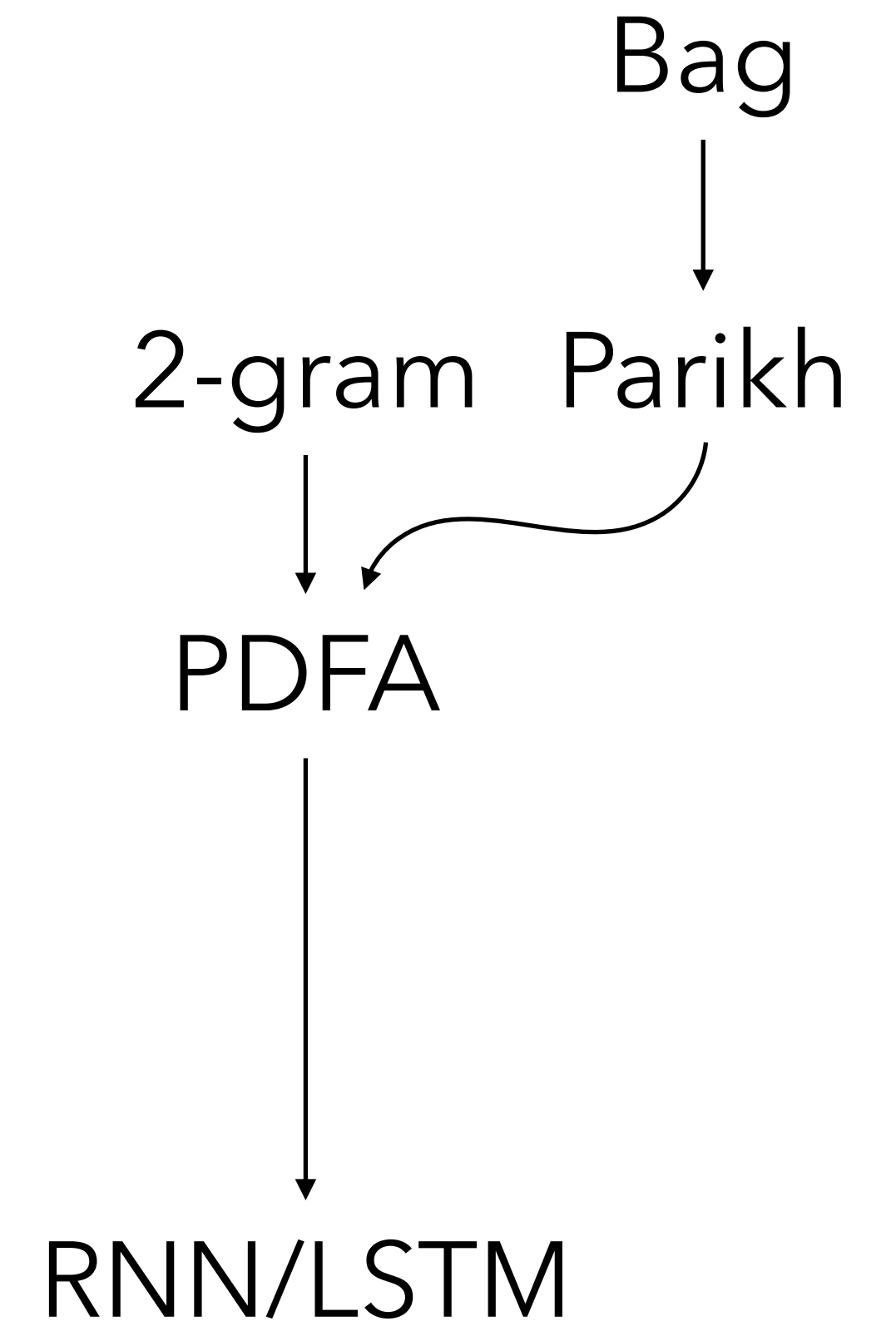
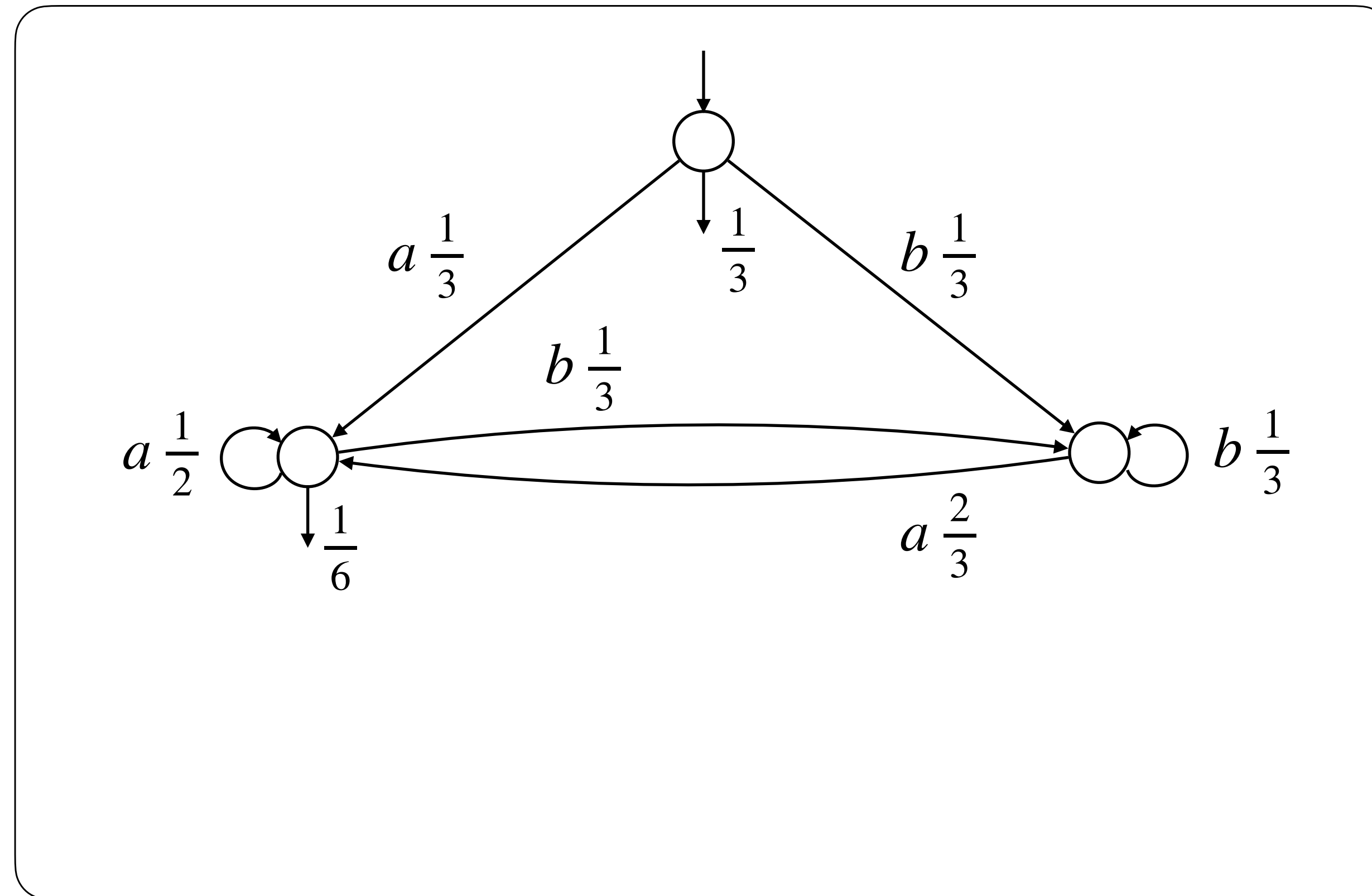


RNN/LSTM

$$p(a \mid a\underline{a}) = p(a \mid b\underline{a}) = \frac{1}{2}$$

Language models

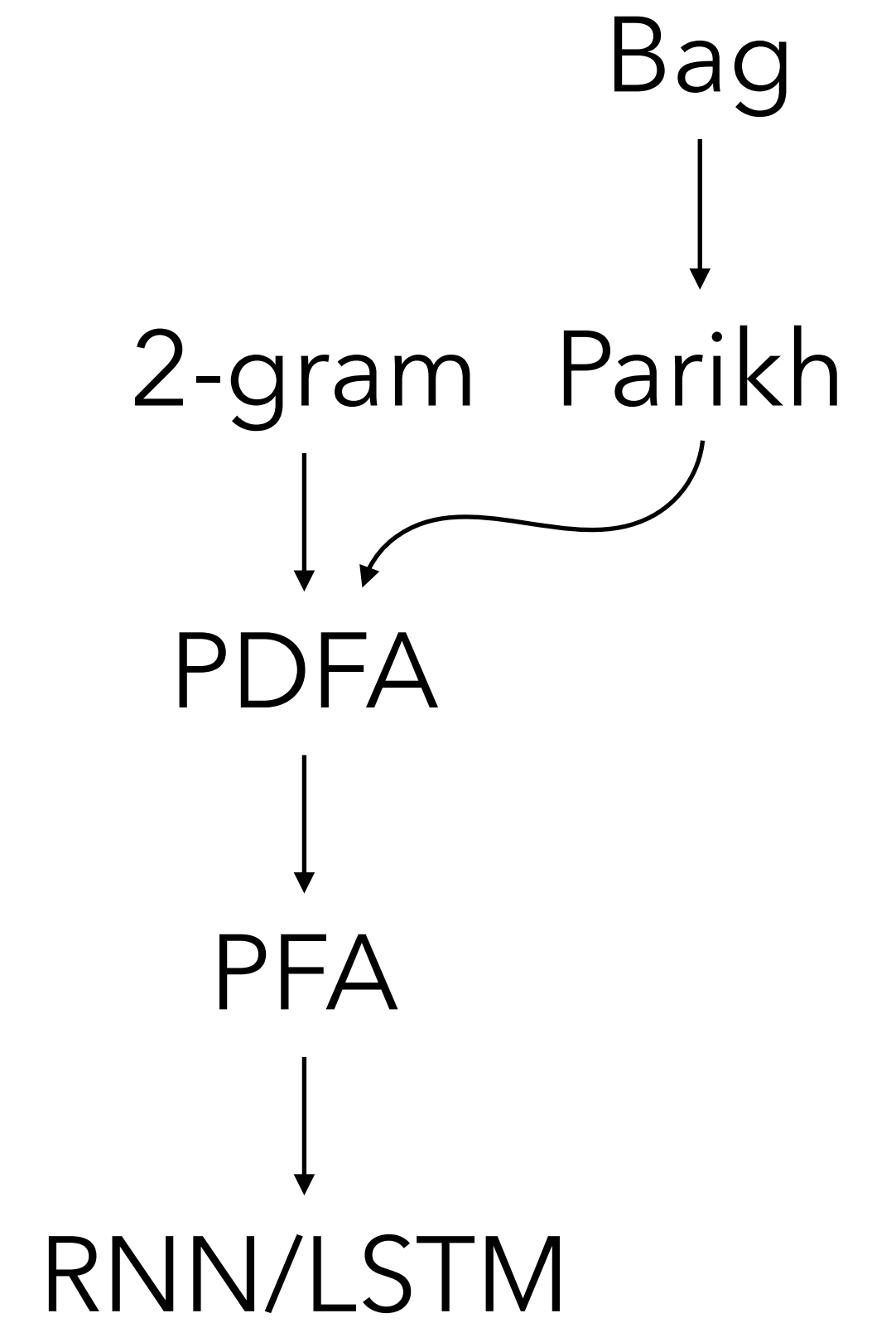
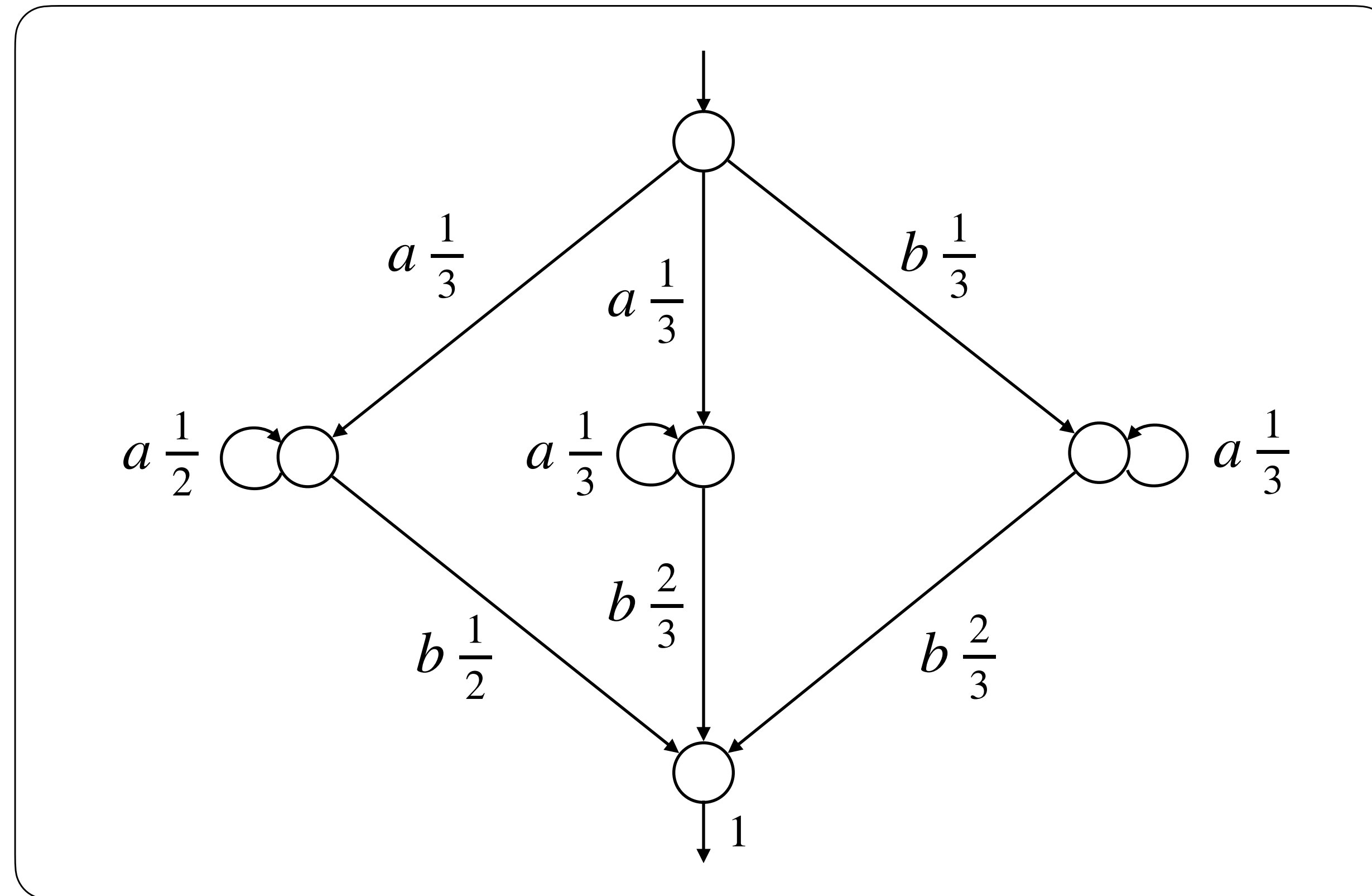
$\Sigma = \{a, b\}$



$$p(a \mid a\underline{a}) = p(a \mid b\underline{a}) = \frac{1}{2}$$

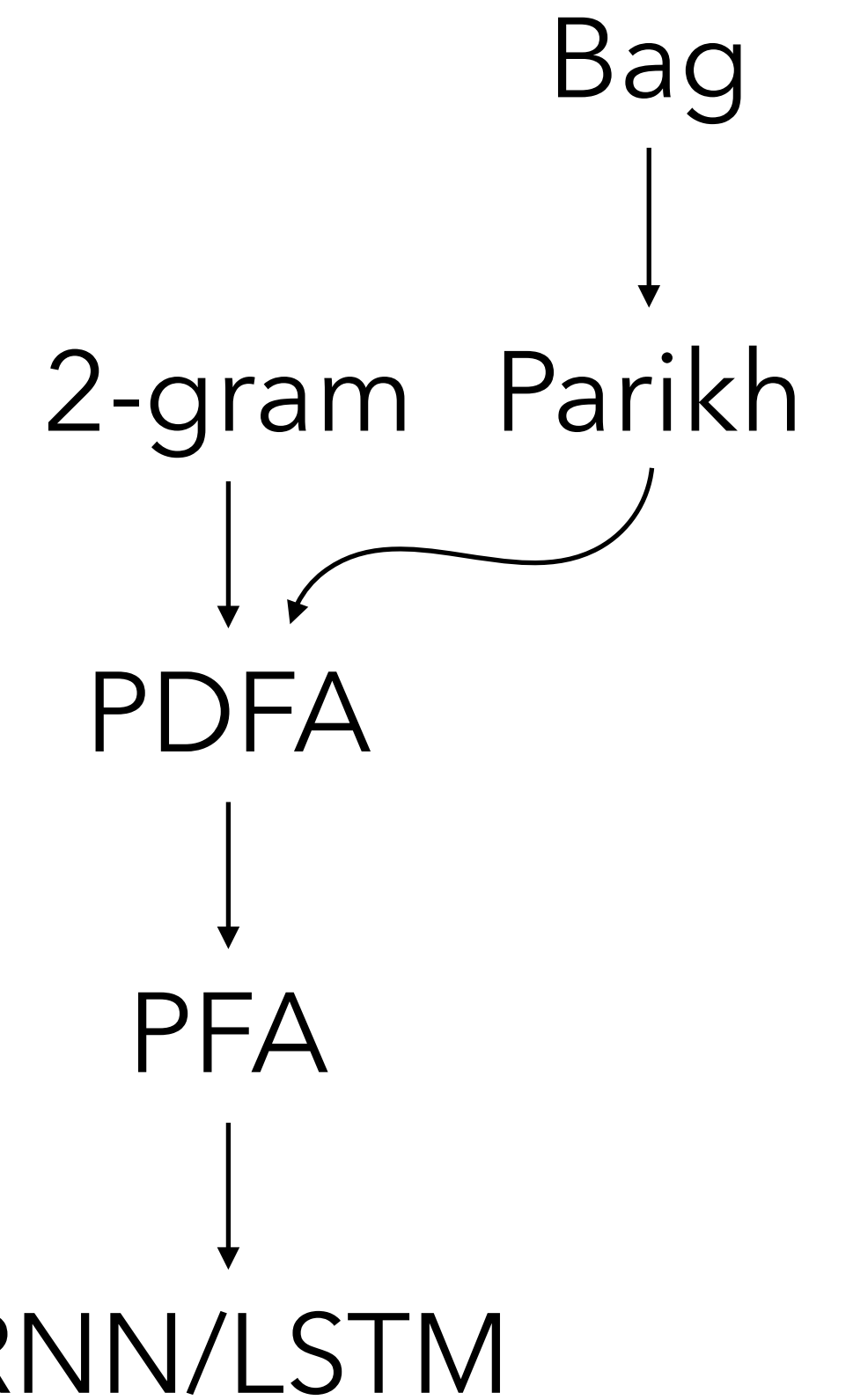
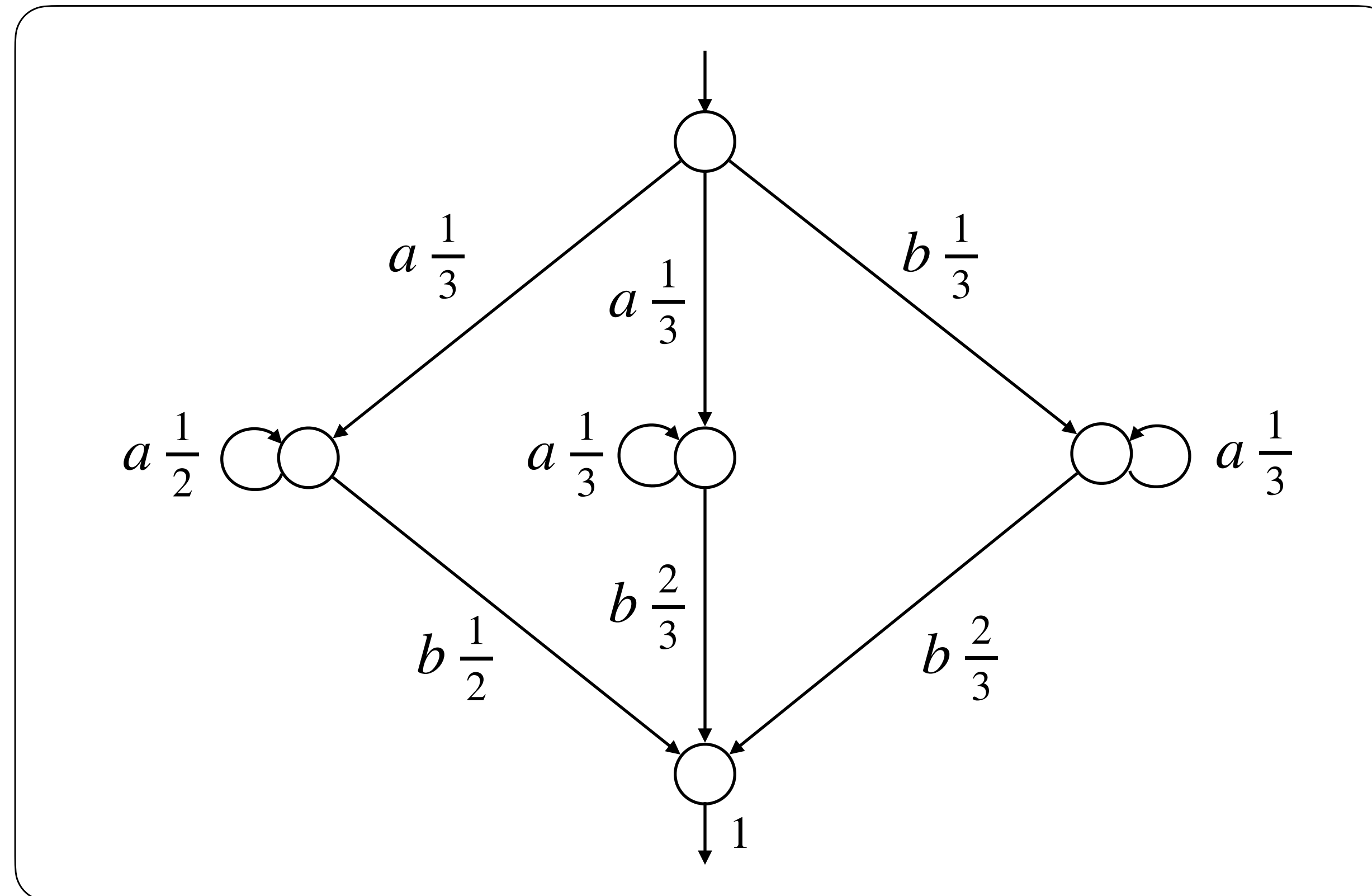
Language models

$\Sigma = \{a, b\}$



Language models

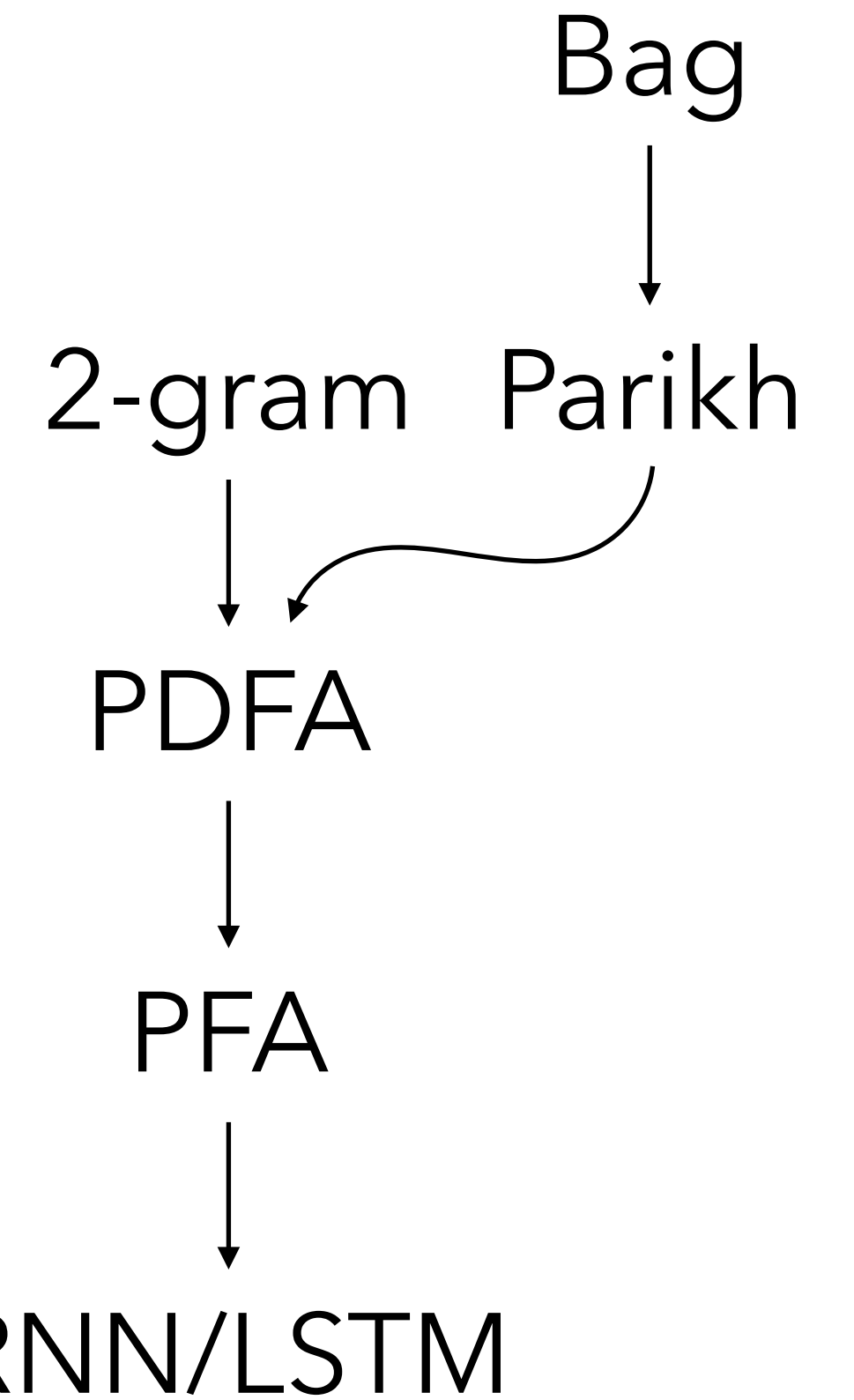
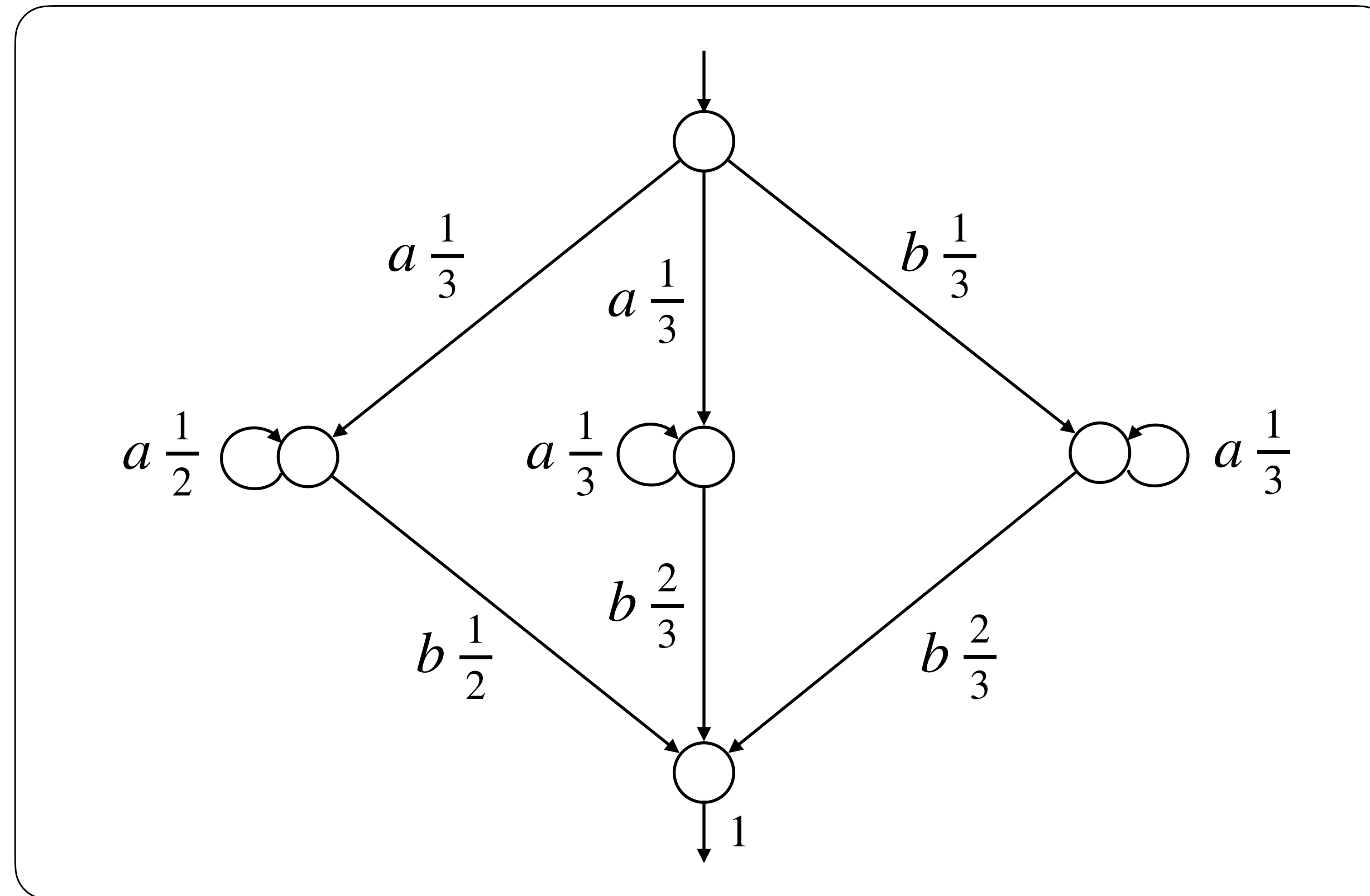
$\Sigma = \{a, b\}$



Svete & Cotterell (2024):
Recurrent neural language
models as probabilistic finite-
state automata.

Language models

$\Sigma = \{a, b\}$



Svete & Cotterell (2024):
Recurrent neural language
models as probabilistic finite-
state automata.

- Slower inference
- Current learning algorithms not suitable for streaming

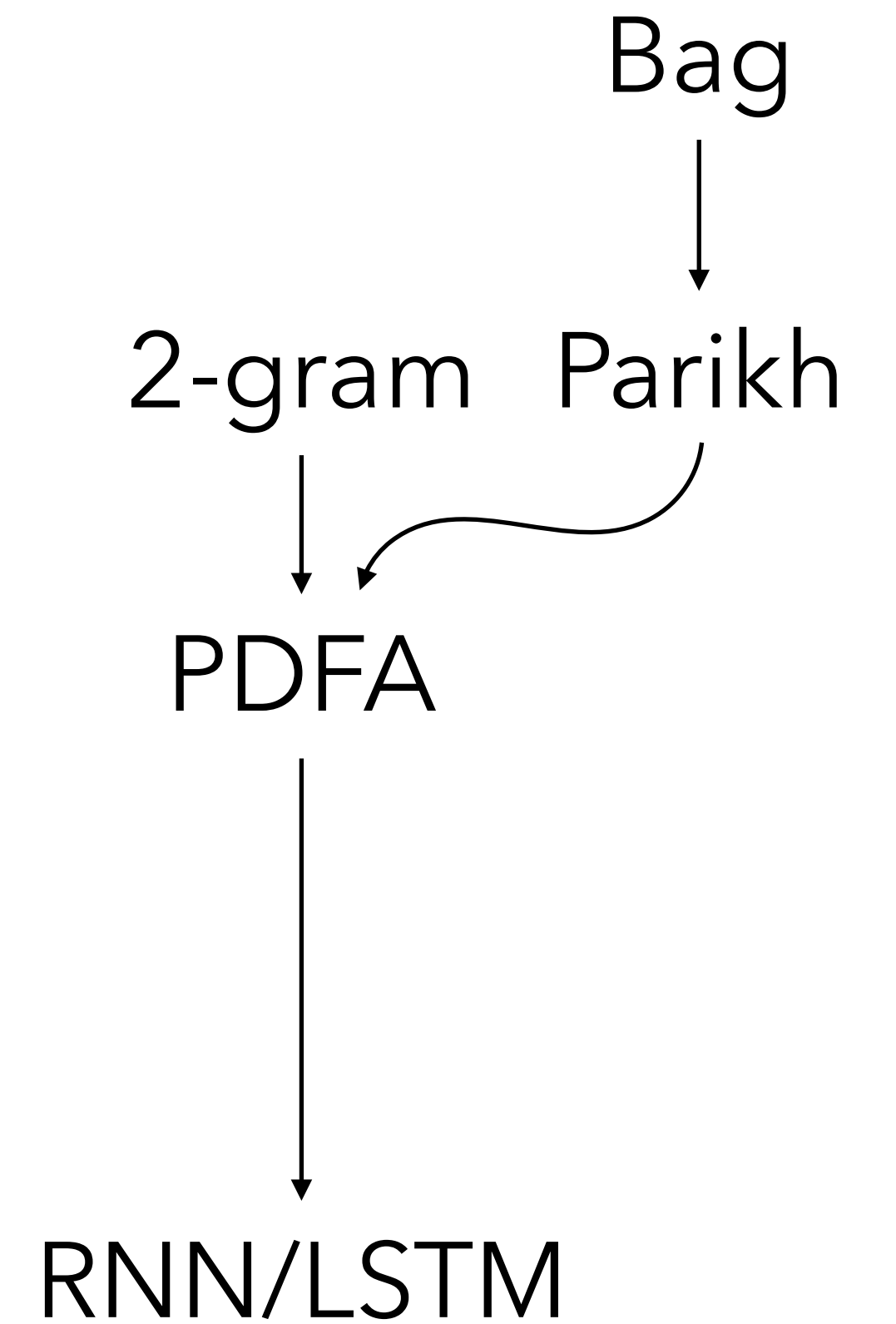
Language models

Σ : finite alphabet of actions

$\Delta(S)$: distributions over S

$$p: \Sigma^* \rightarrow \Delta(\Sigma \cup \{\text{stop}\})$$

$$\begin{cases} Q \rightarrow \Delta(\Sigma \cup \{\text{stop}\}) \\ Q \times \Sigma \rightarrow Q \end{cases} \quad \text{incremental}$$



Language models

Σ : finite alphabet of actions

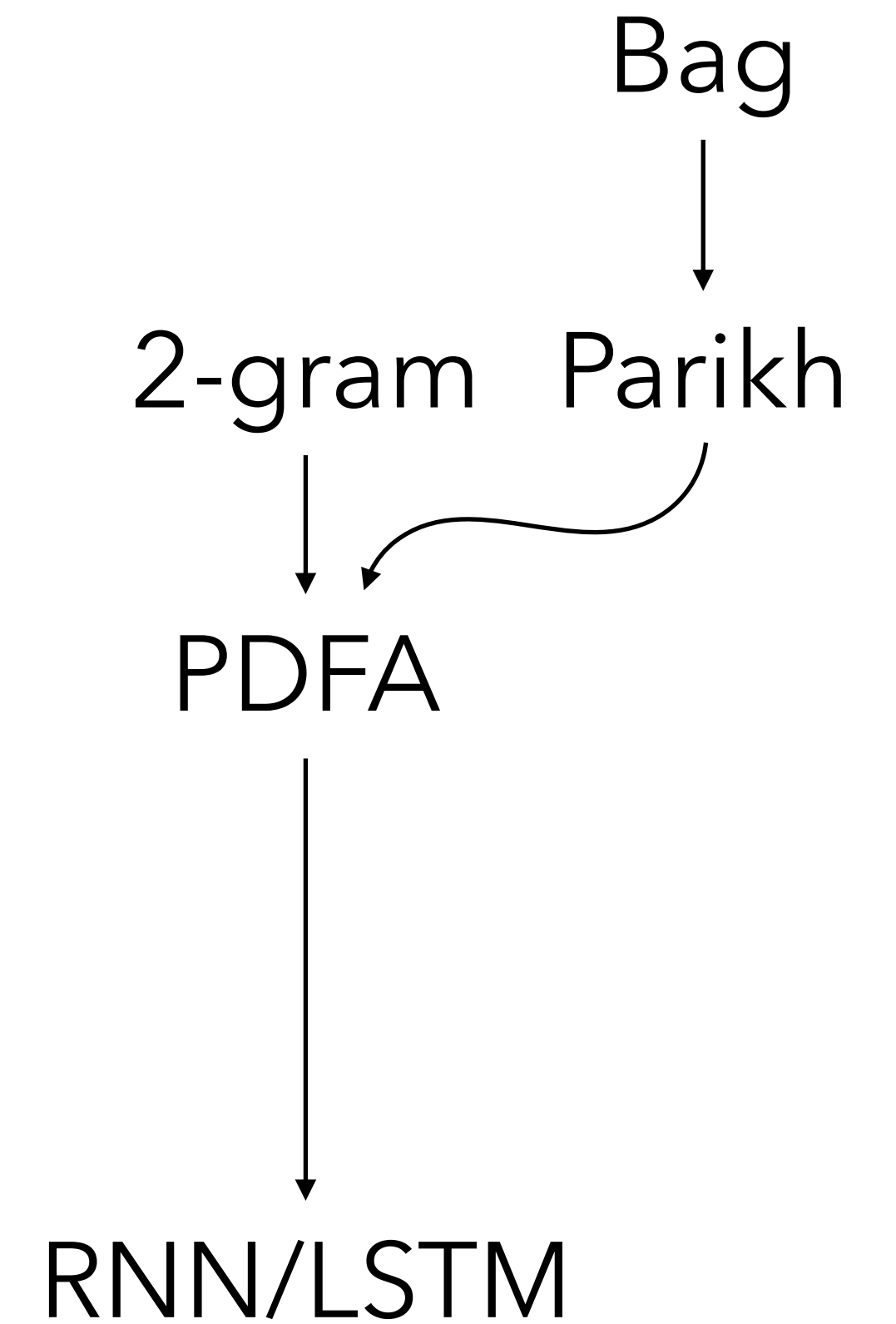
$\Delta(S)$: distributions over S

$$p: \Sigma^* \rightarrow \Delta(\Sigma \cup \{\text{stop}\})$$

state reached after sequence

$$\begin{cases} Q \rightarrow \Delta(\Sigma \cup \{\text{stop}\}) \\ Q \times \Sigma \rightarrow Q \end{cases}$$

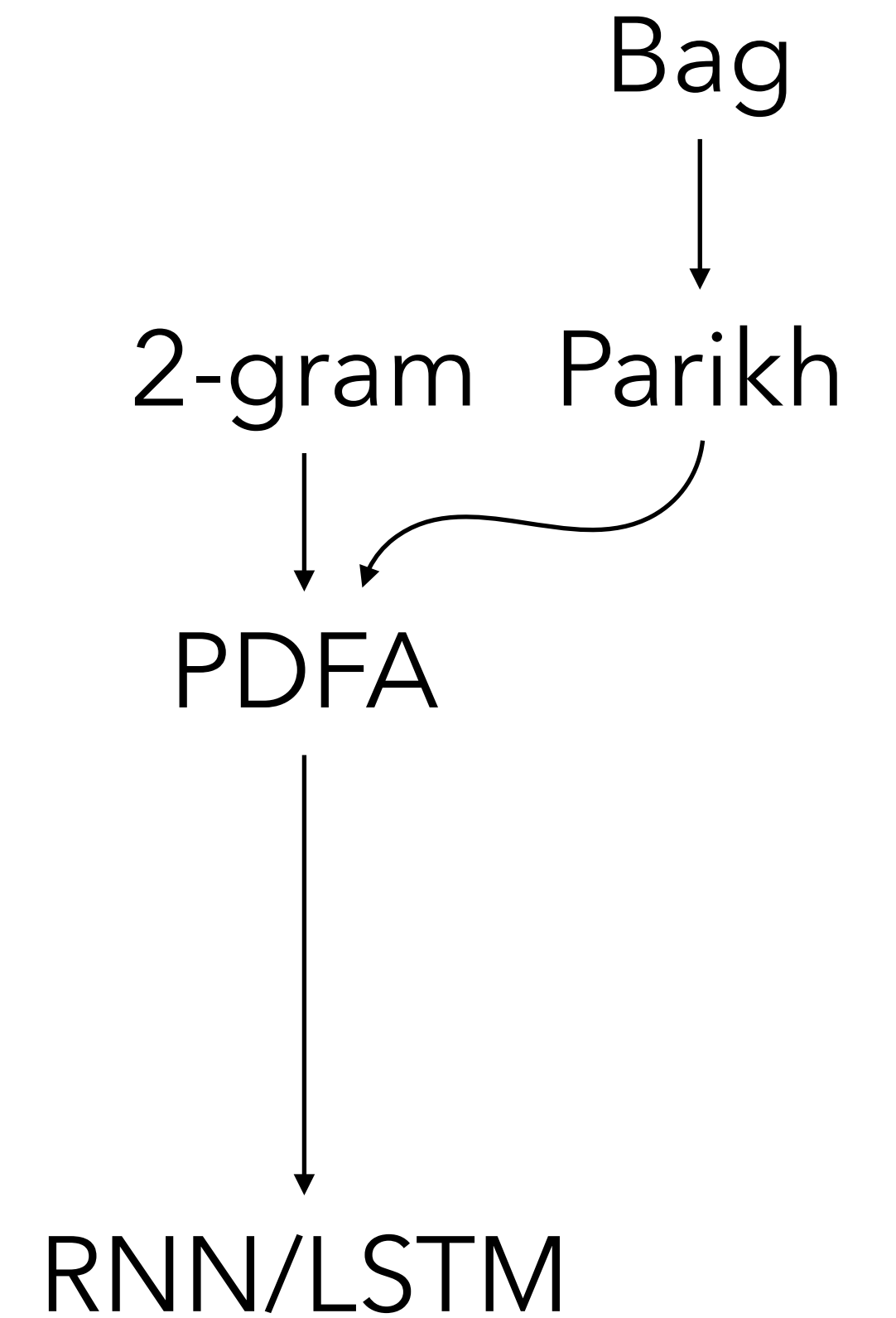
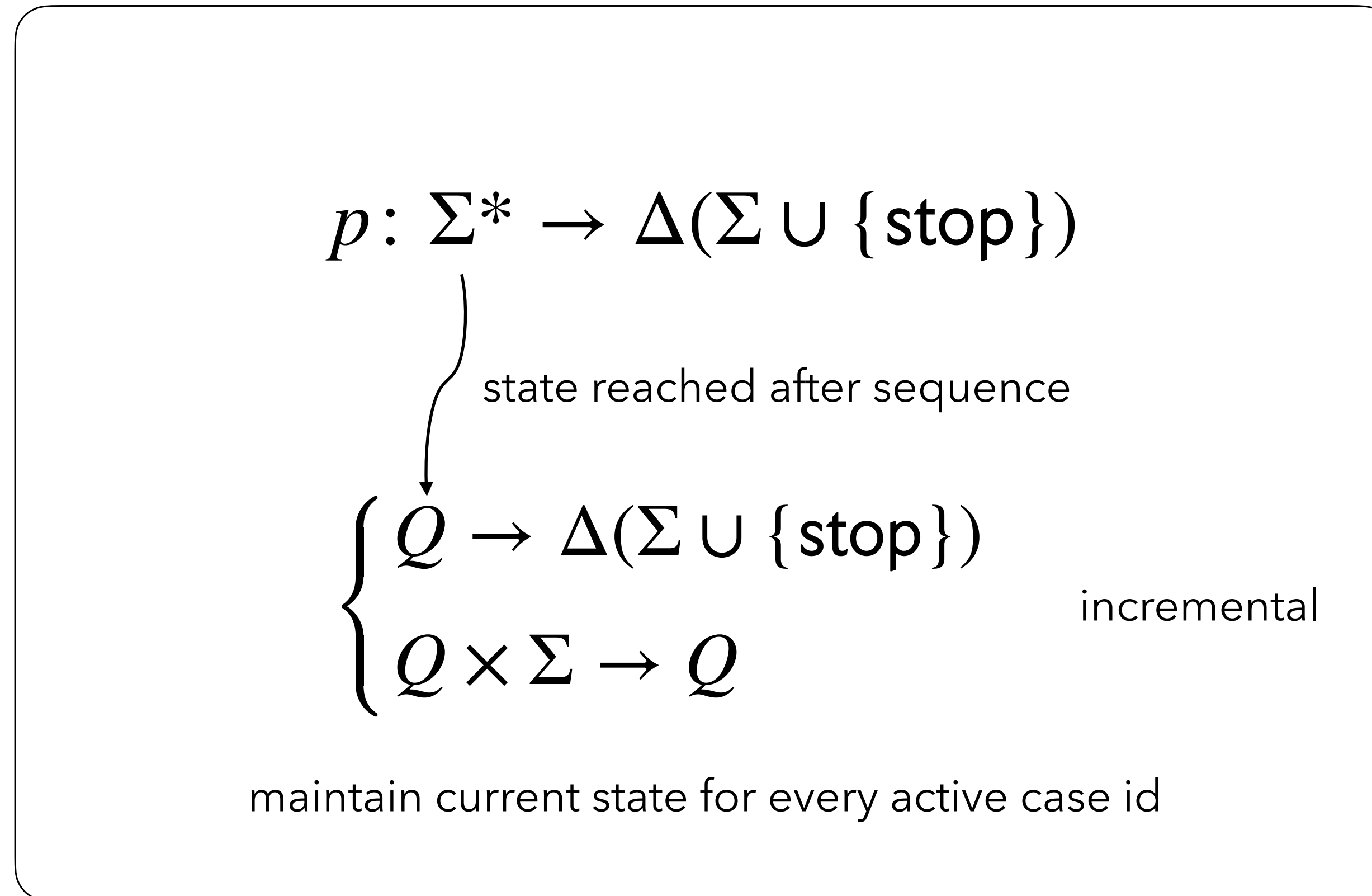
incremental



Language models

Σ : finite alphabet of actions

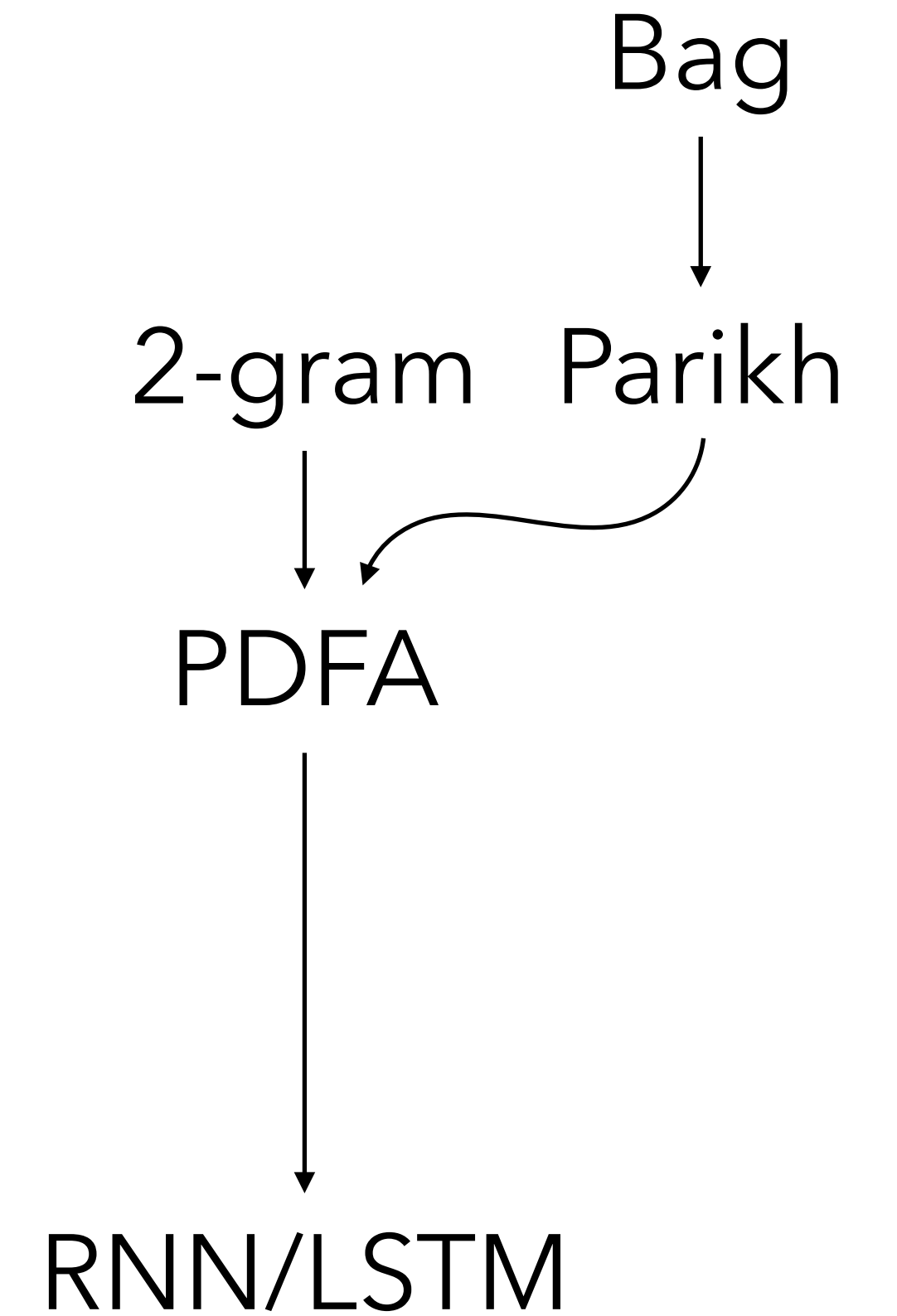
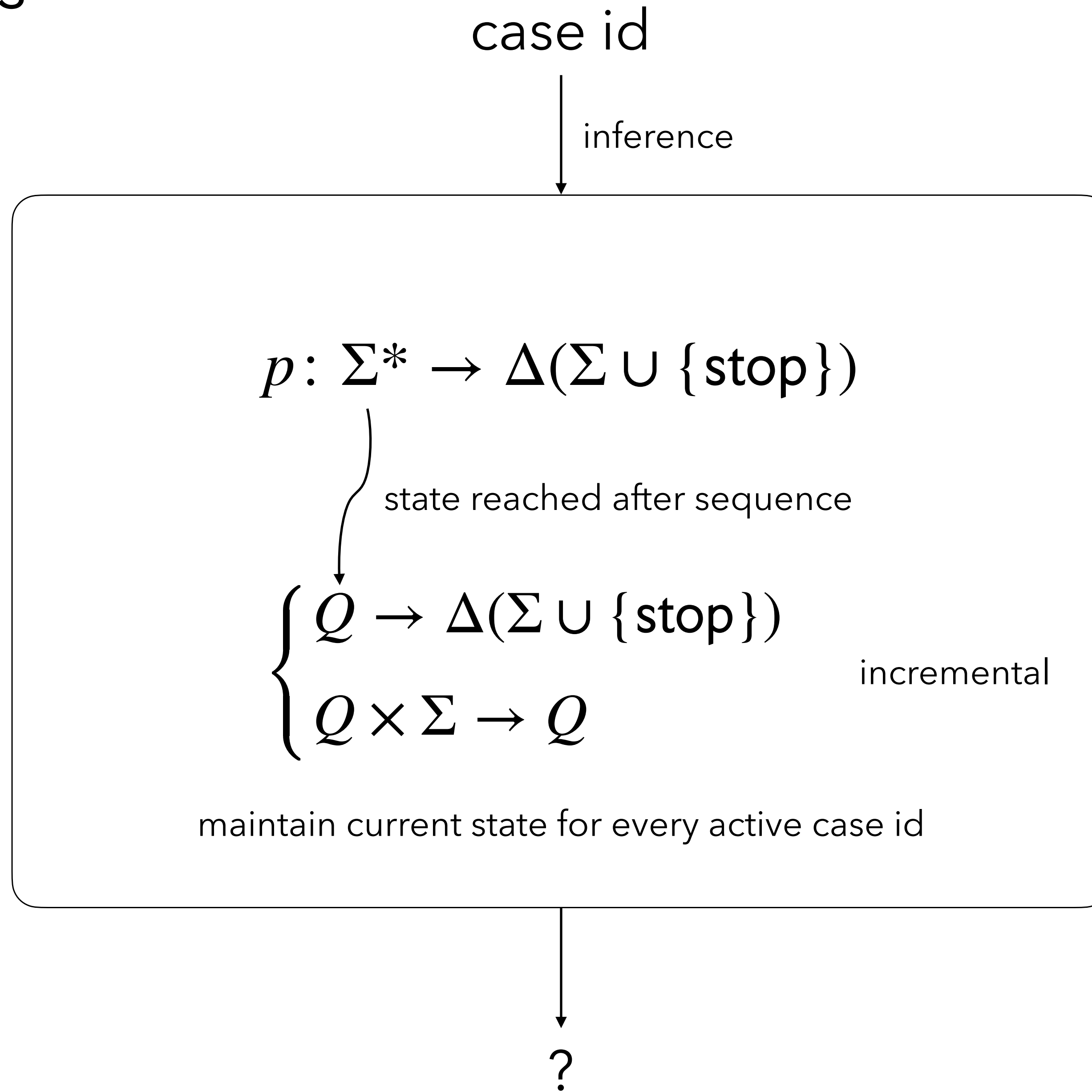
$\Delta(S)$: distributions over S



Language models

Σ : finite alphabet of actions

$\Delta(S)$: distributions over S



Batch Learning

Batch learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

Batch learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



Batch learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	?

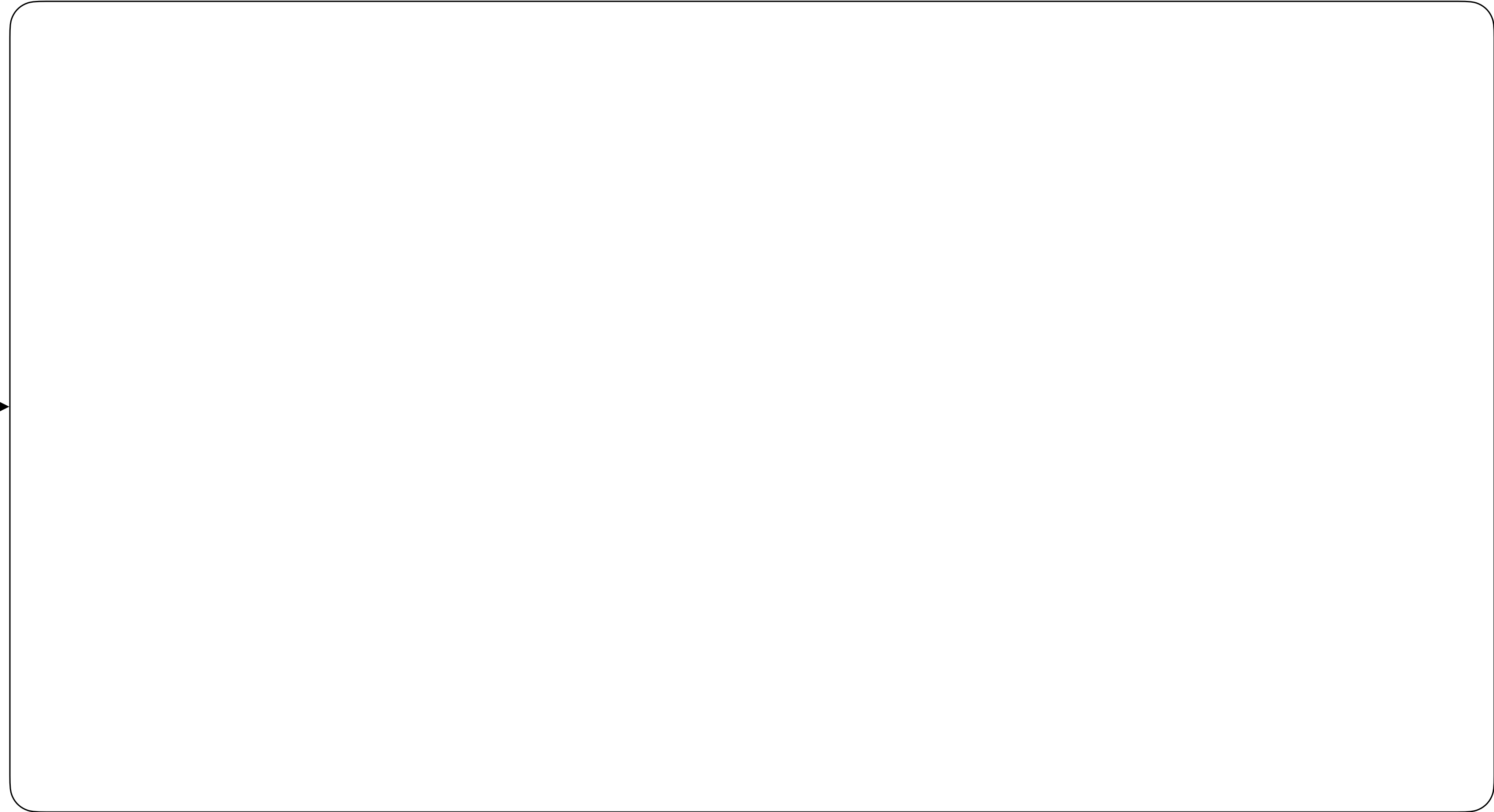
train



Batch learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



Batch learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

enough to

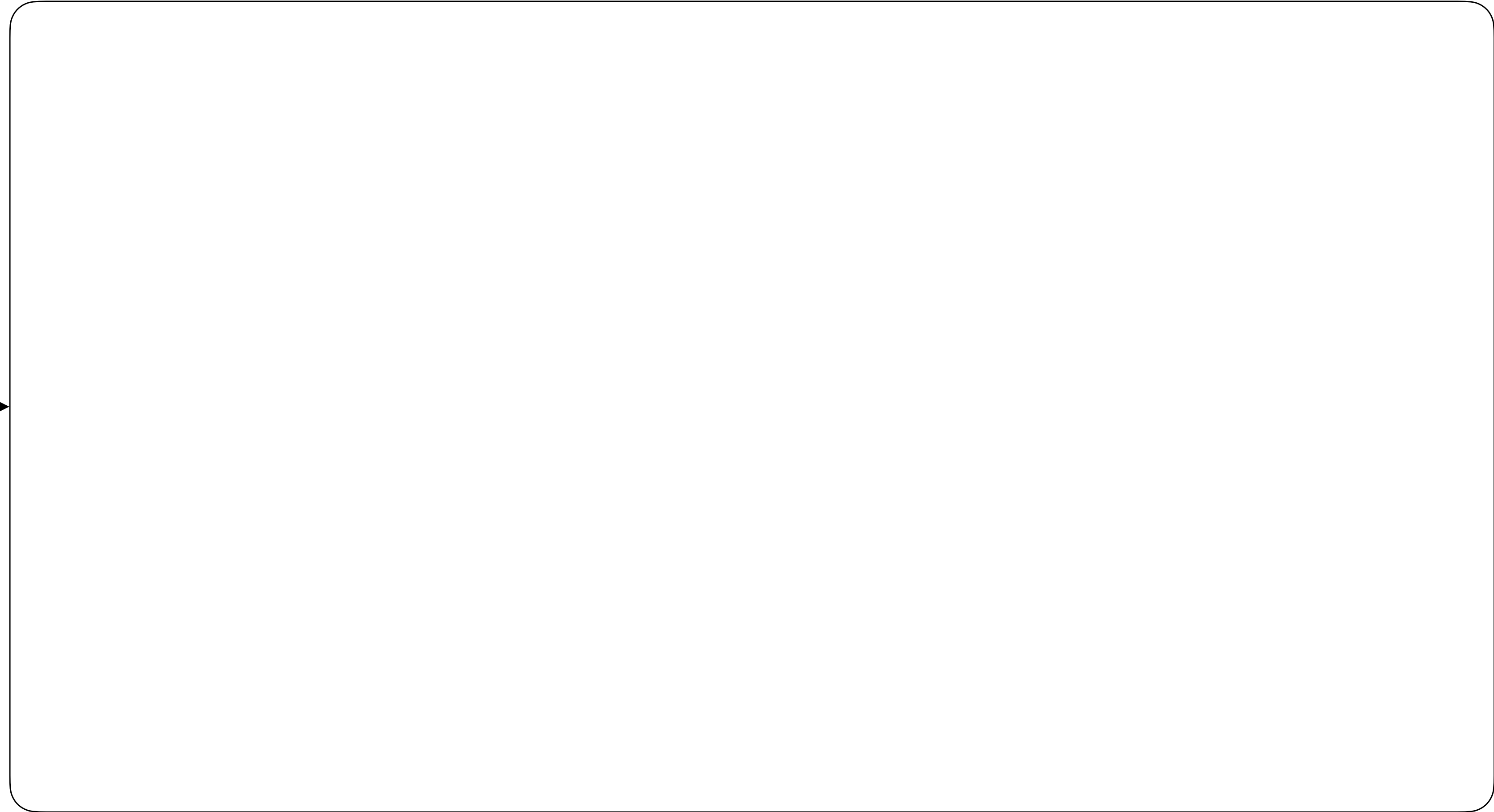
maintain event order

within each case

event-log

$$L : \Sigma^* \rightarrow \mathbb{N}$$

train



Batch learning

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	$\langle bbb \rangle^1$ CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



Batch learning of RNN/LSTM

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	$\langle bbb \rangle^1$ CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train

backpropagation through time
(BPTT)

Batch learning of PDFFA

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	$\langle bbb \rangle^1$ CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

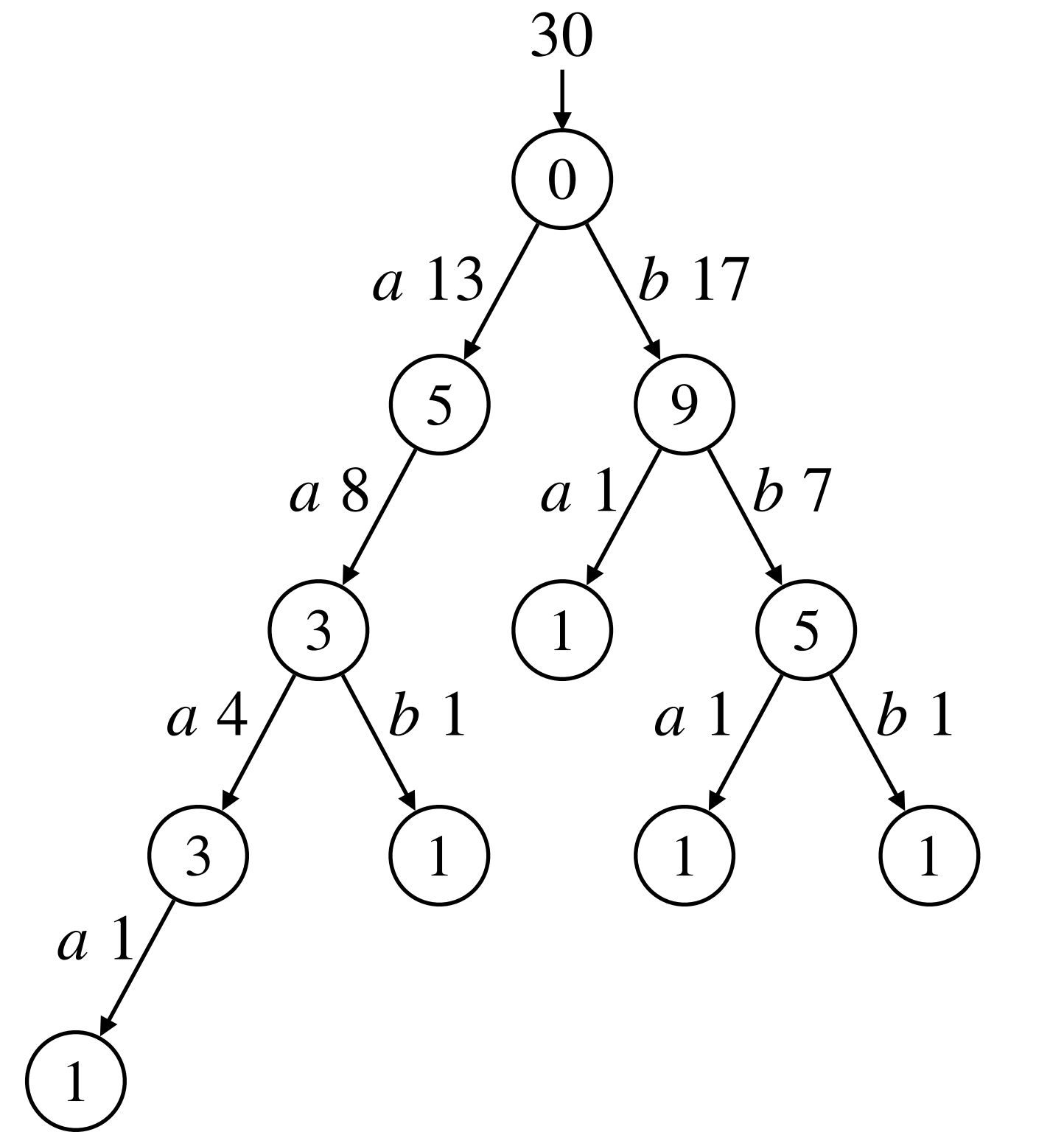
train



Batch learning of PDFFA

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	$\langle aaaa \rangle^1$ ER Registration
I	ER Triage
I	$\langle b \rangle^9$ ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



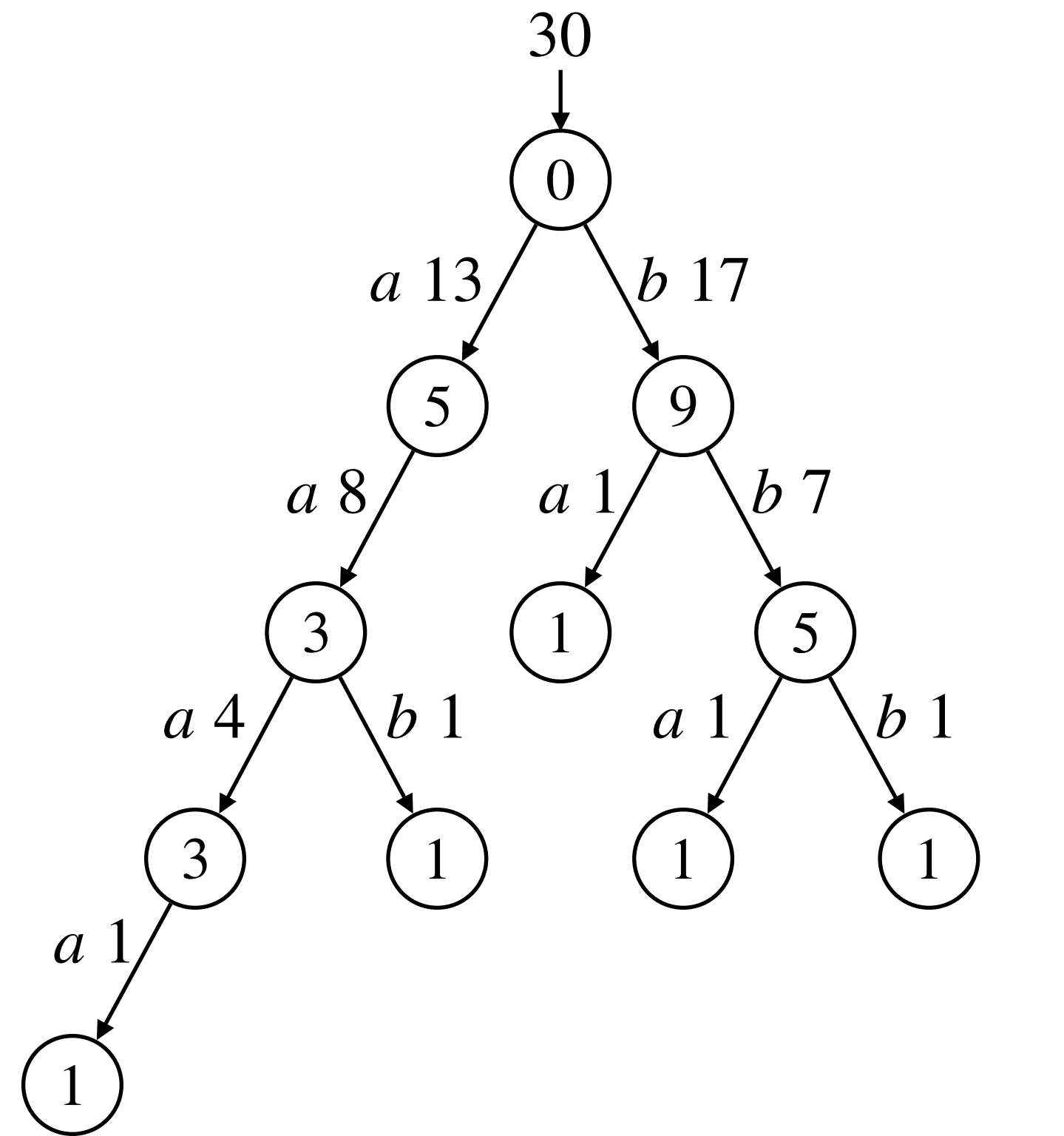
frequency prefix-tree acceptor (FPTA)

Batch learning of PDFFA

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	$\langle aaaa \rangle^1$ ER Registration
I	ER Triage
I	$\langle b \rangle^9$ ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train

exploit frequencies



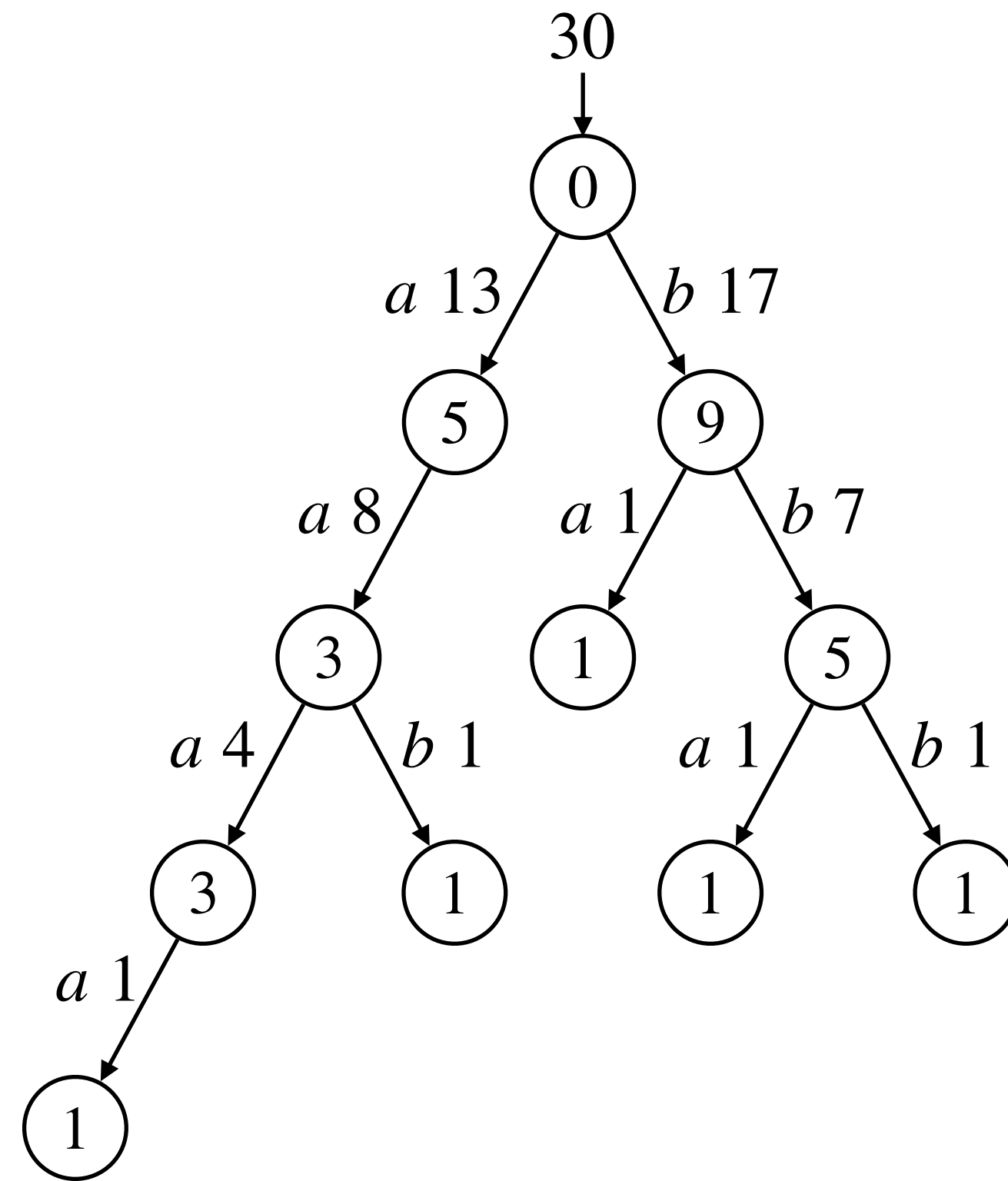
frequency prefix-tree acceptor (FPTA)

Batch learning of PDFFA

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	$\langle aaaa \rangle^1$ ER Registration
I	ER Triage
I	$\langle b \rangle^9$ ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

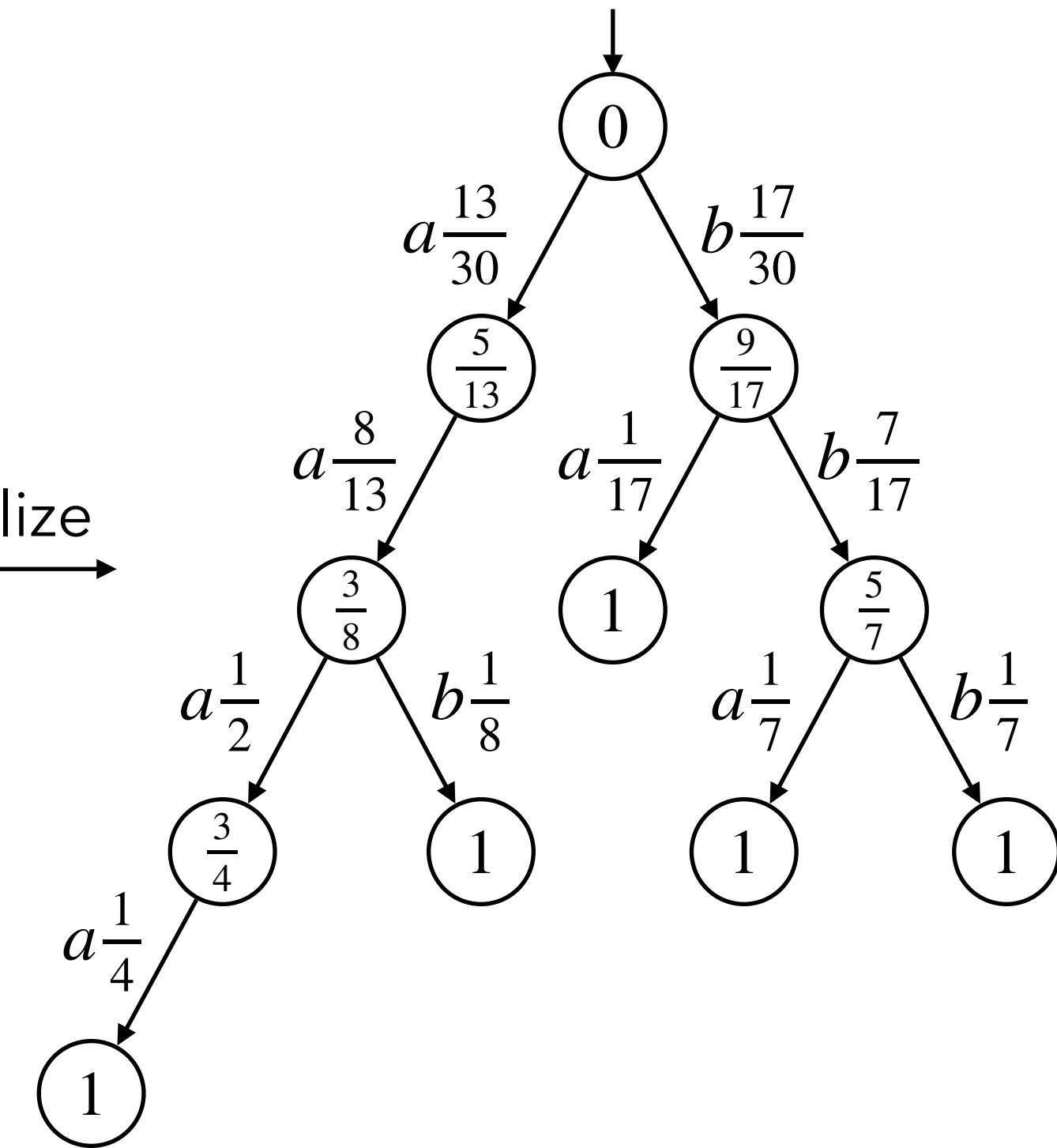
train

exploit frequencies



frequency prefix-tree acceptor (FPTA)

normalize

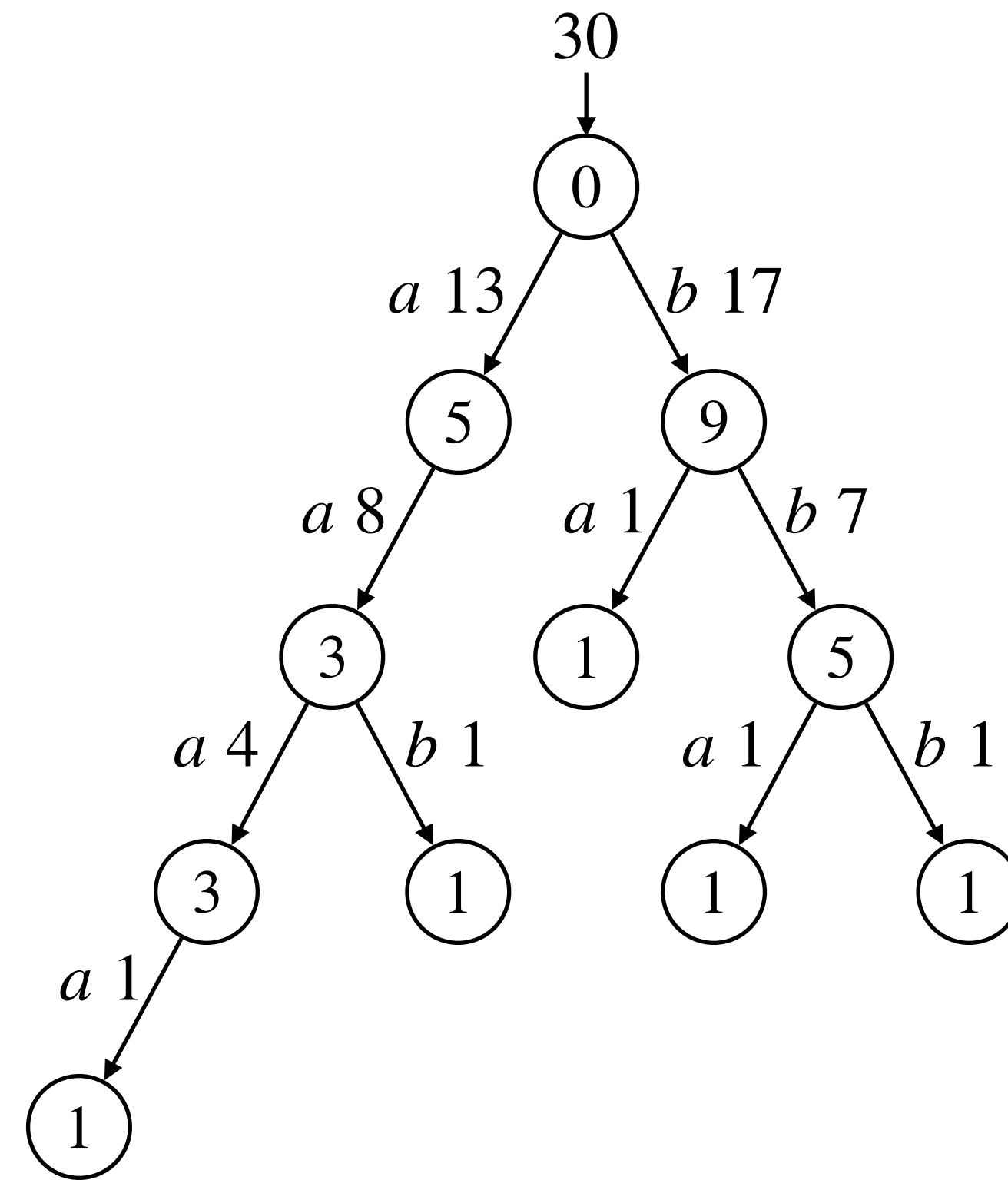


tree-shaped PDFFA

Batch learning of PDFFA

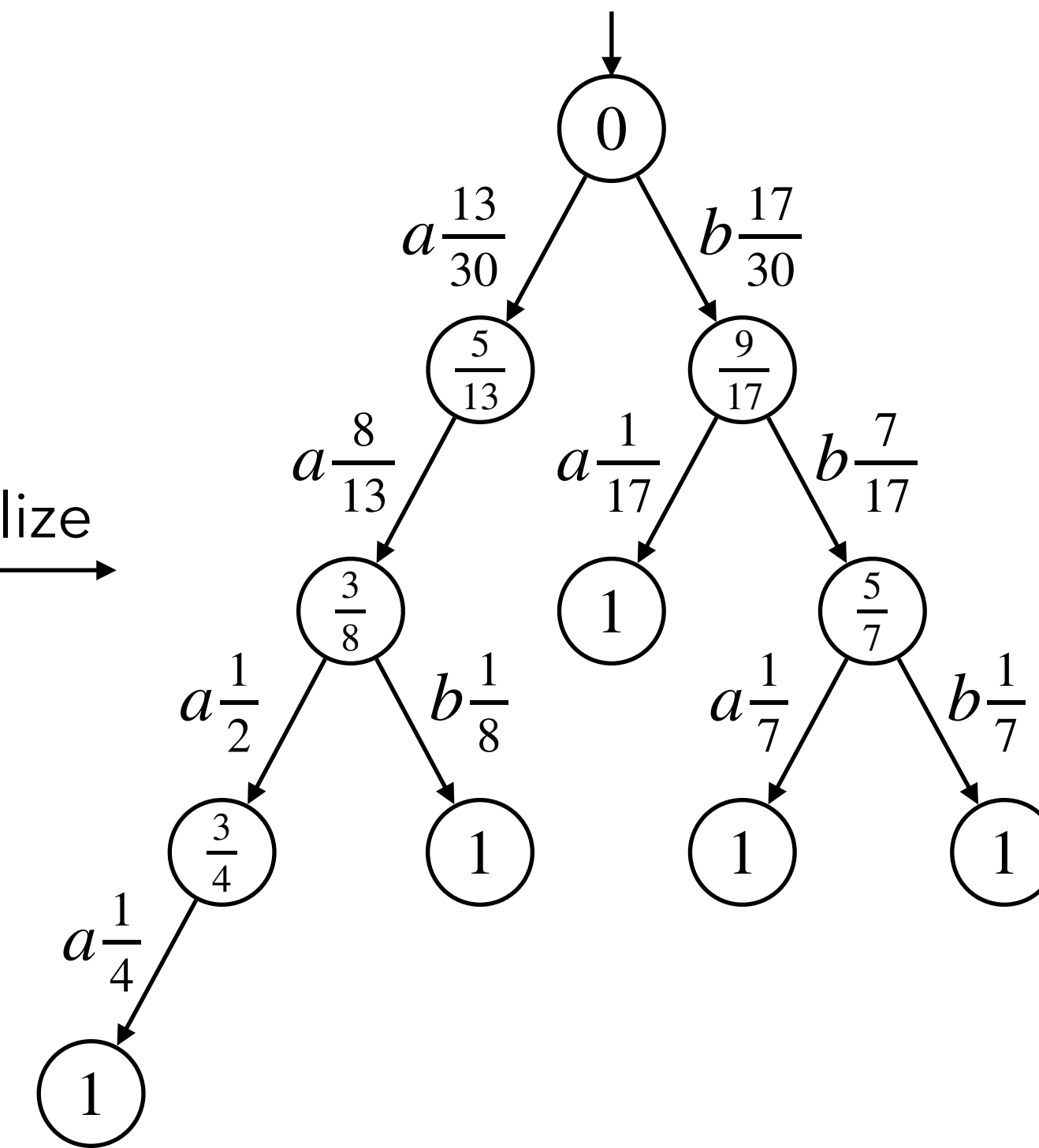
case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	$\langle aaaa \rangle^1$ ER Registration
I	ER Triage
I	$\langle b \rangle^9$ ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train
exploit frequencies



frequency prefix-tree acceptor (FPTA)

normalize



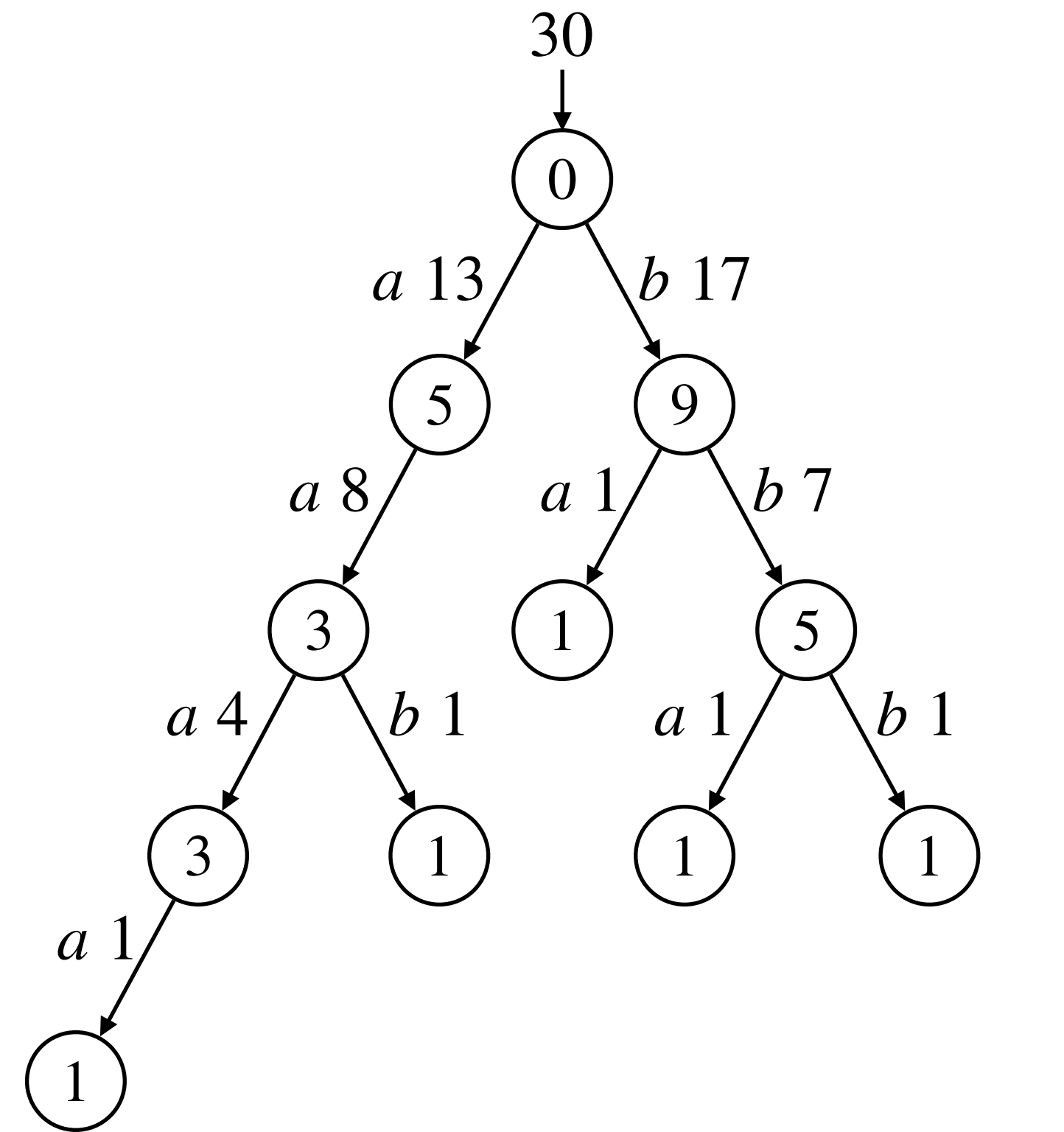
tree-shaped PDFFA

as a predictor, does not generalize well
many sequences cannot be parsed

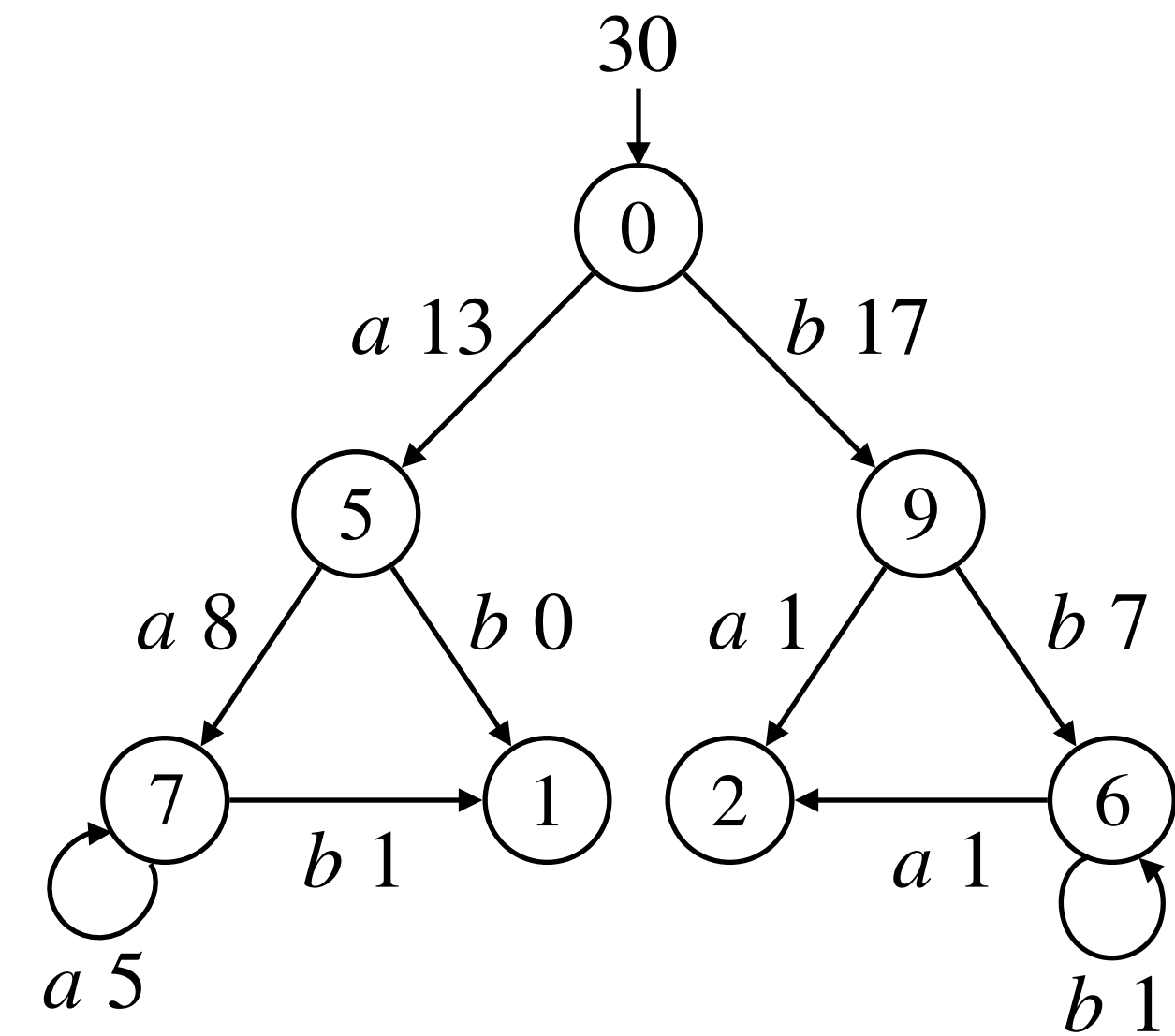
Batch learning of PDFFA

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	$\langle aaaa \rangle^1$ ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



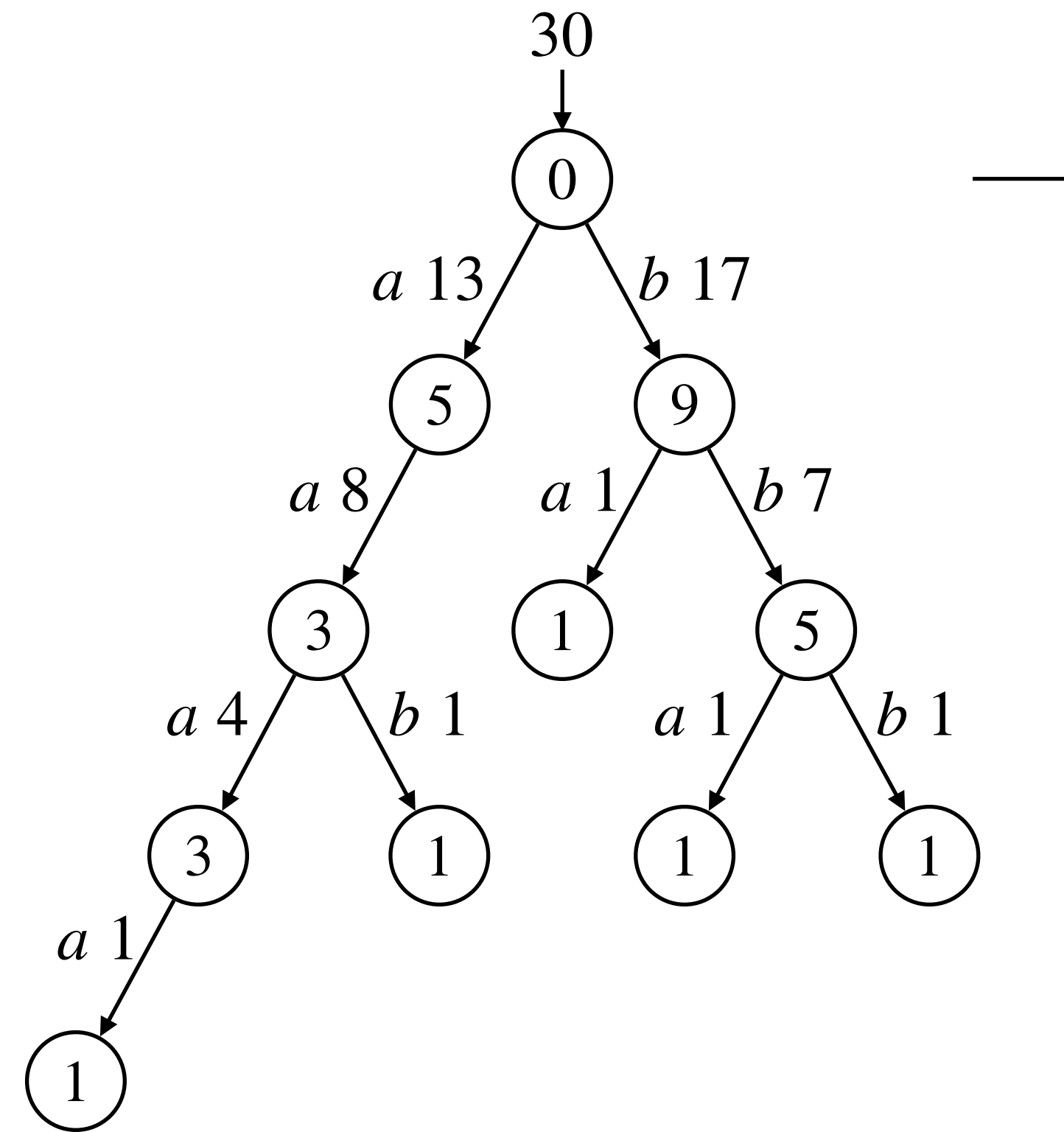
frequency prefix-tree acceptor (FPTA)



Batch learning of PDFFA

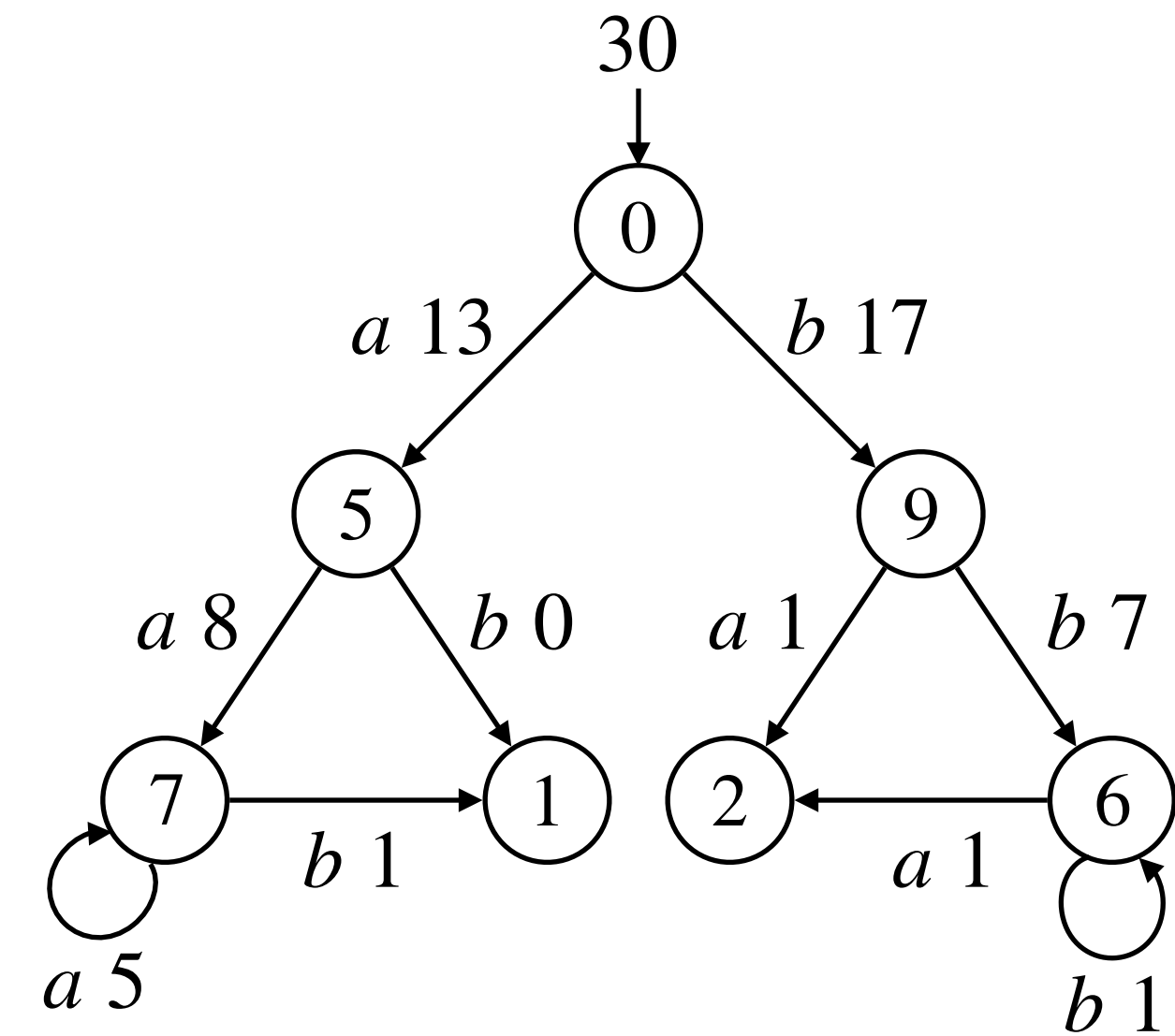
case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



frequency prefix-tree acceptor (FPTA)

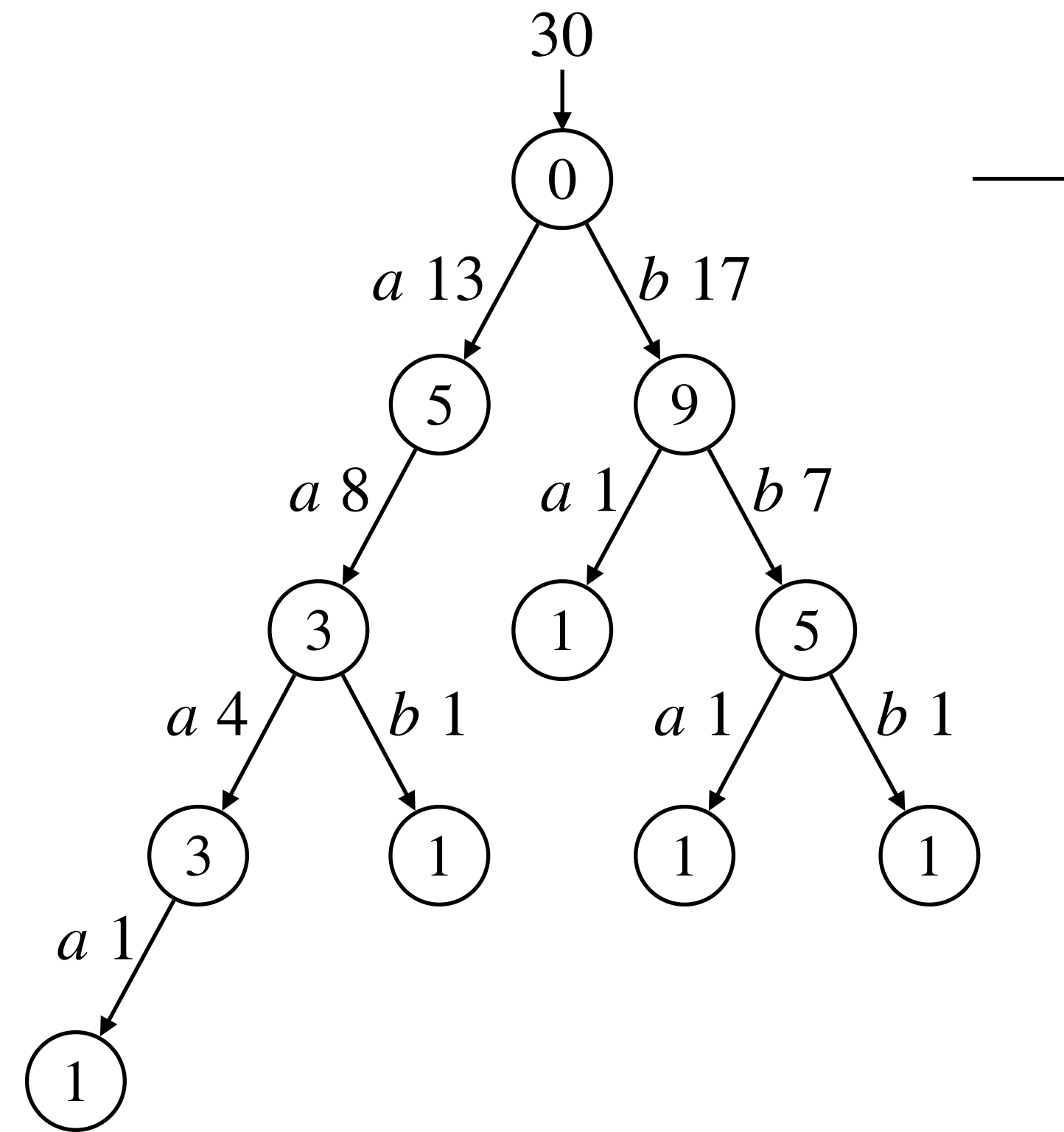
fold



Batch learning of PDFFA

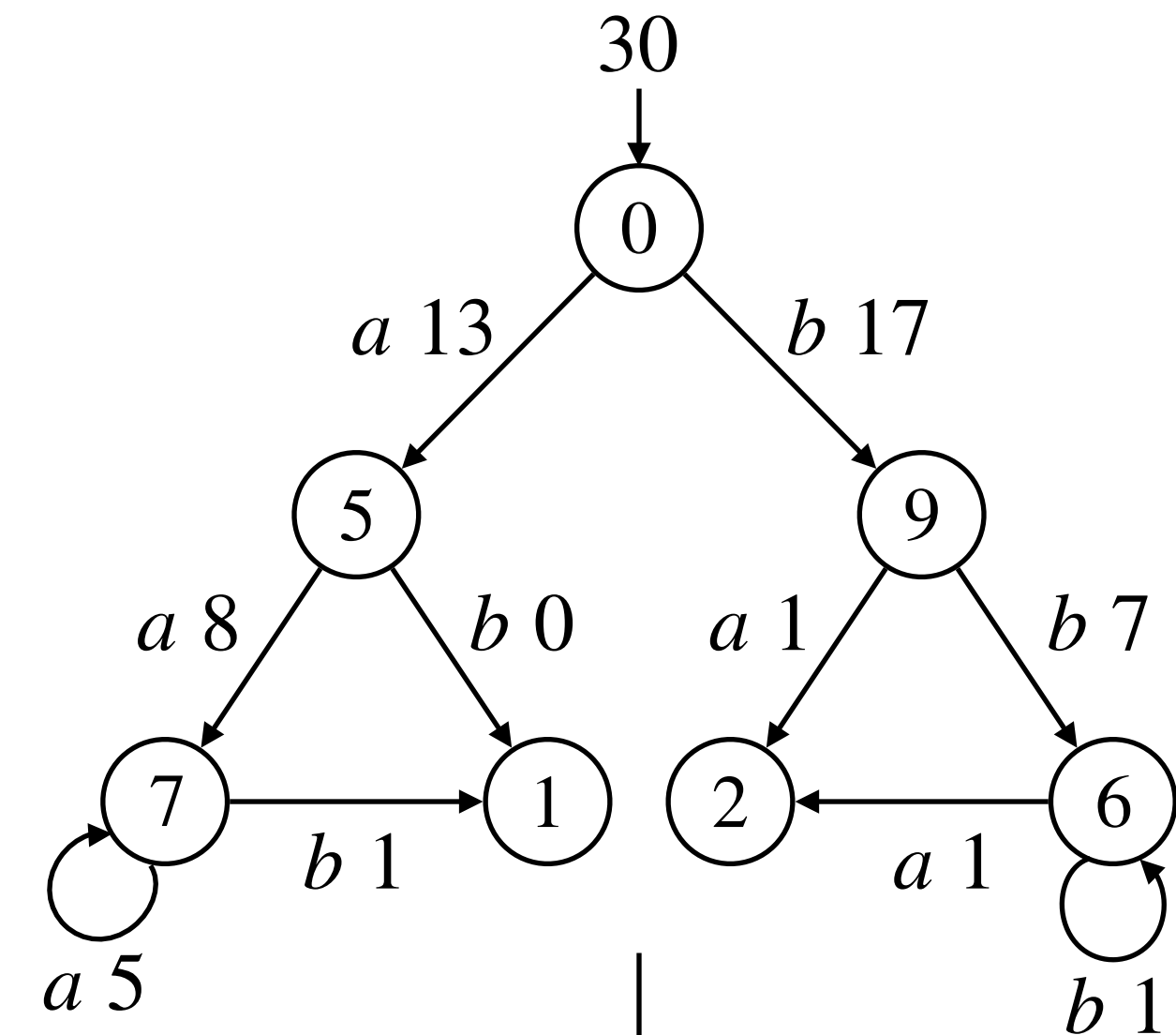
case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



frequency prefix-tree acceptor (FPTA)

fold



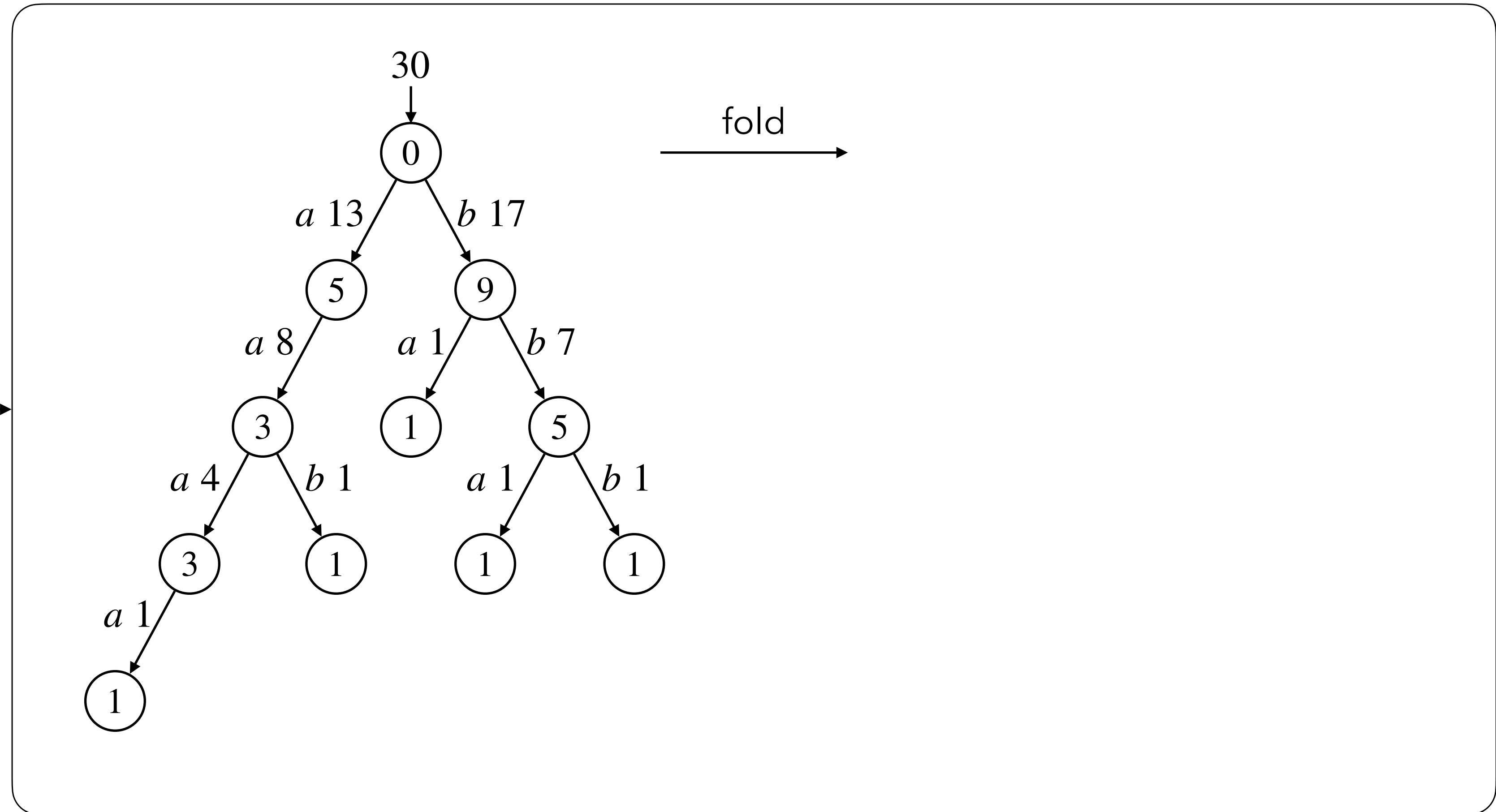
normalize

PDFFA

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	$\langle bbb \rangle^1$ CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train

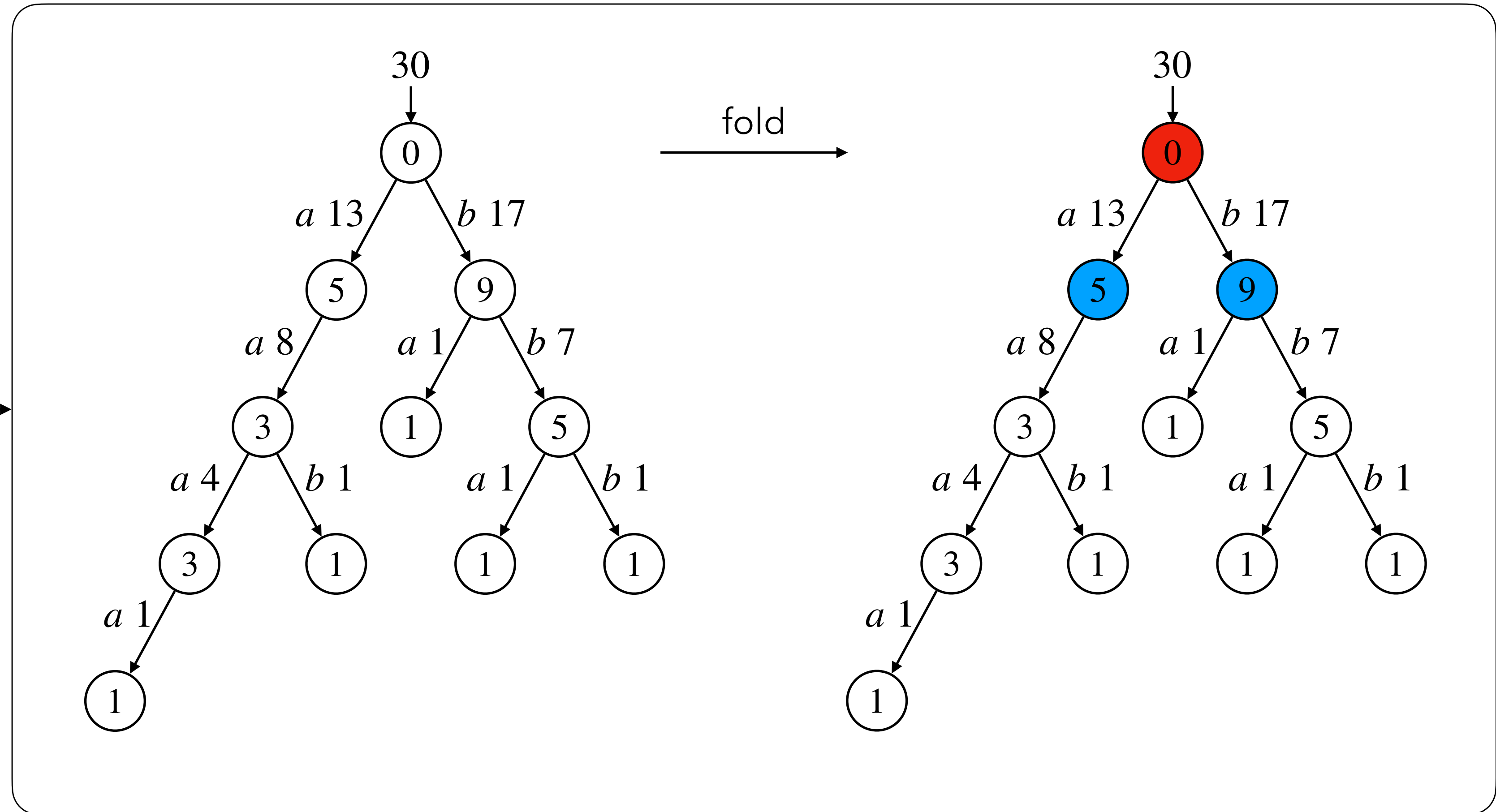


Carrasco & Oncina (1994): Learning stochastic regular grammars by means of a state merging method.

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	$\langle bb \rangle^5$ Leucocytes
I	Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



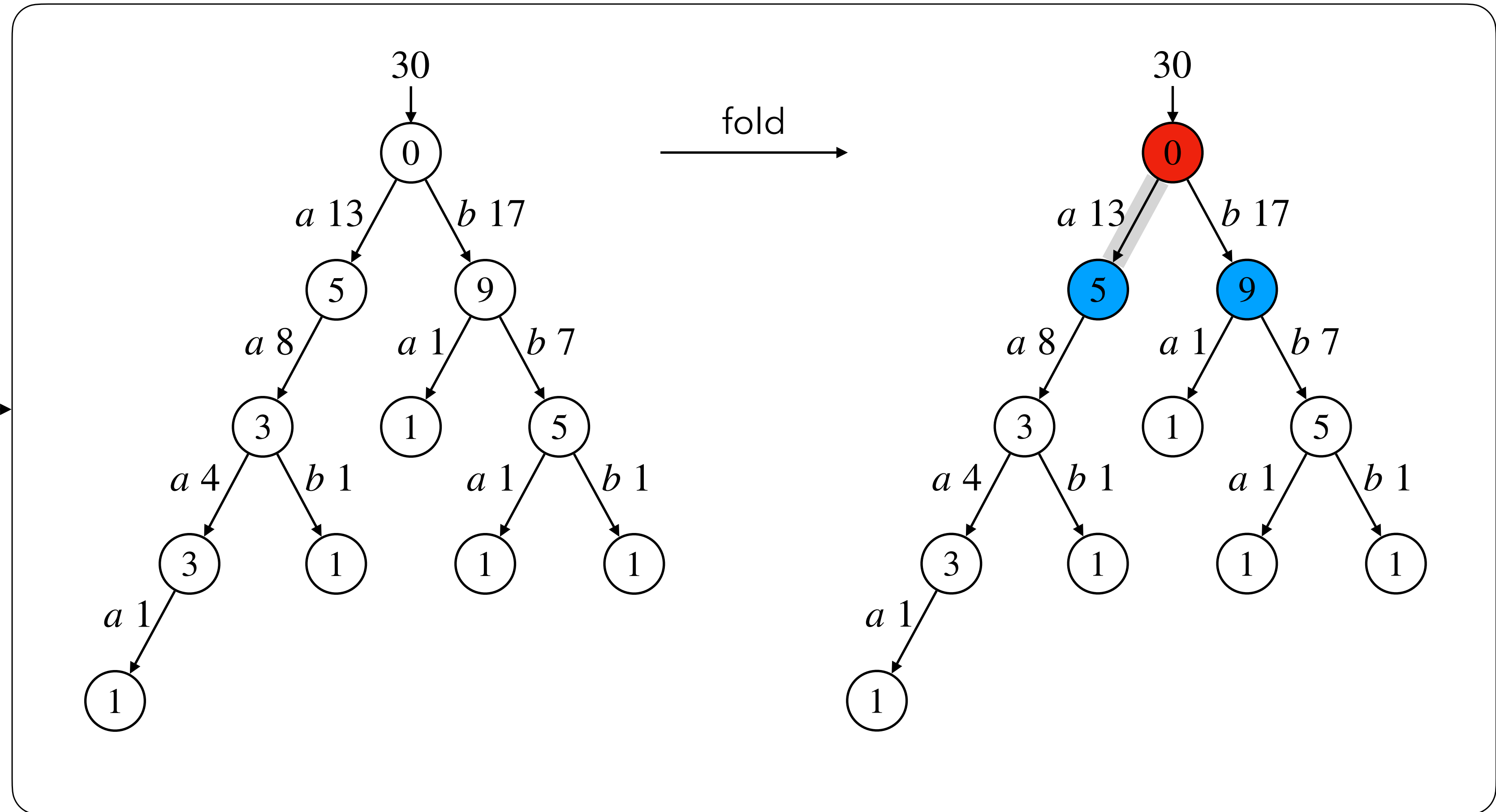
- Top-down method

- Merges **blue states** into **red states**

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



- Top-down method

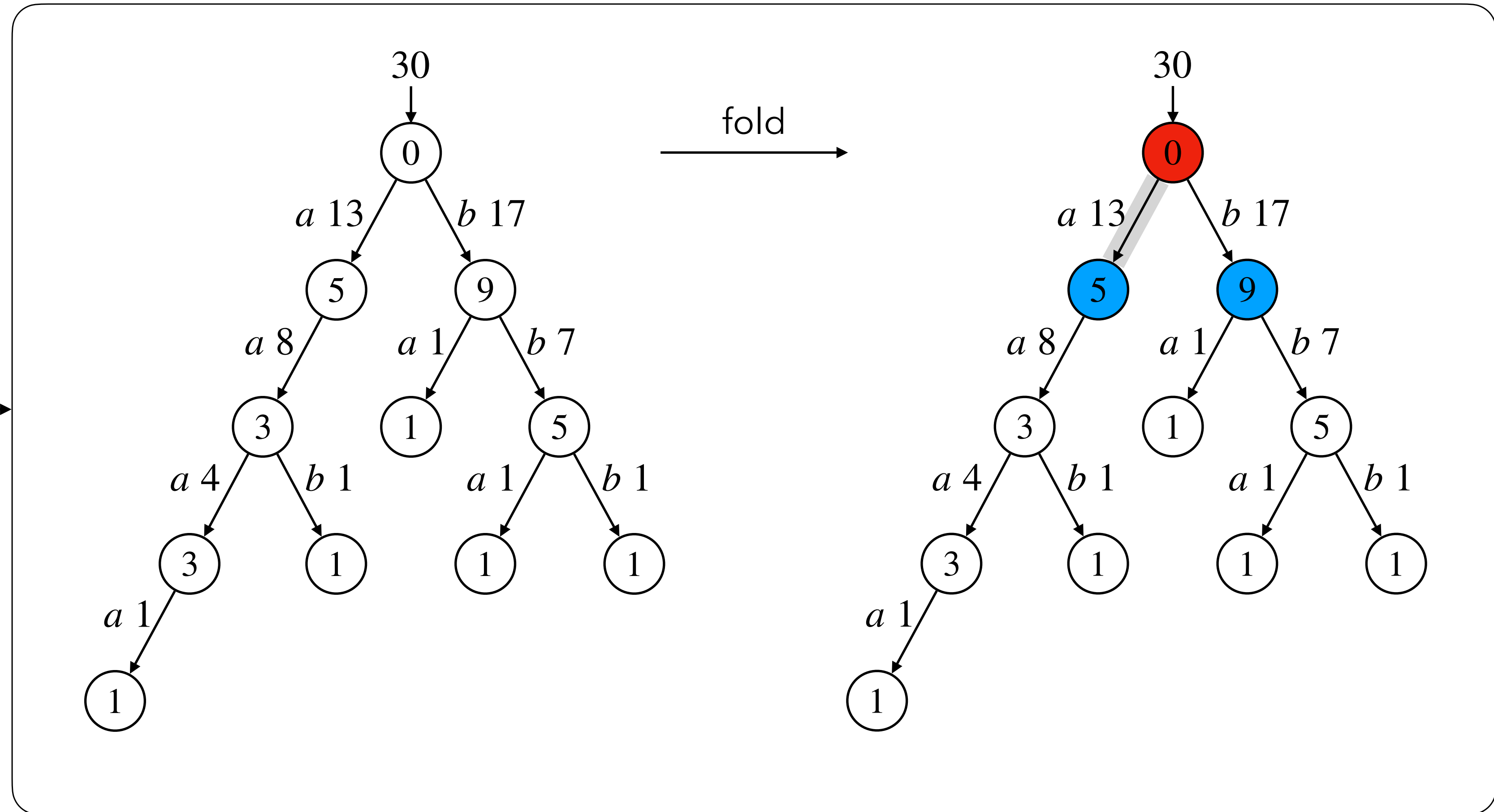
- Merges **blue states** into **red states**

Batch learning of PDFFA: Alergia

do not merge if their probabilities are statistically different

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	$\langle bb \rangle^5$ Leucocytes
I	Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	$\langle bbb \rangle^1$ CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



- Top-down method

- Merges **blue states** into **red states**

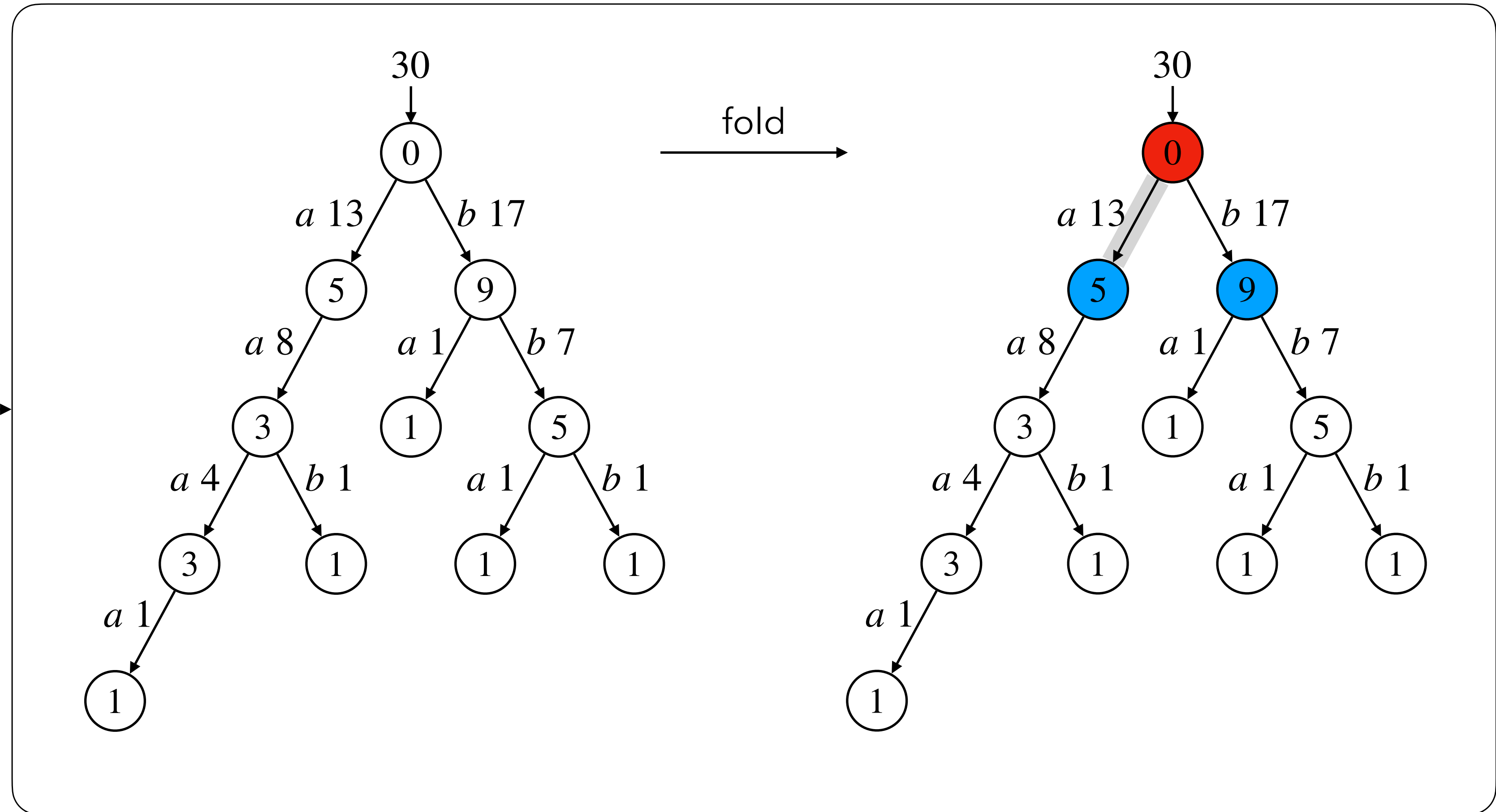
Batch learning of PDFFA: Alergia

do not merge if their probabilities are statistically different

probabilities to stop: $\frac{5}{13} \stackrel{?}{\approx} \frac{0}{30}$

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



- Top-down method

- Merges **blue states** into **red states**

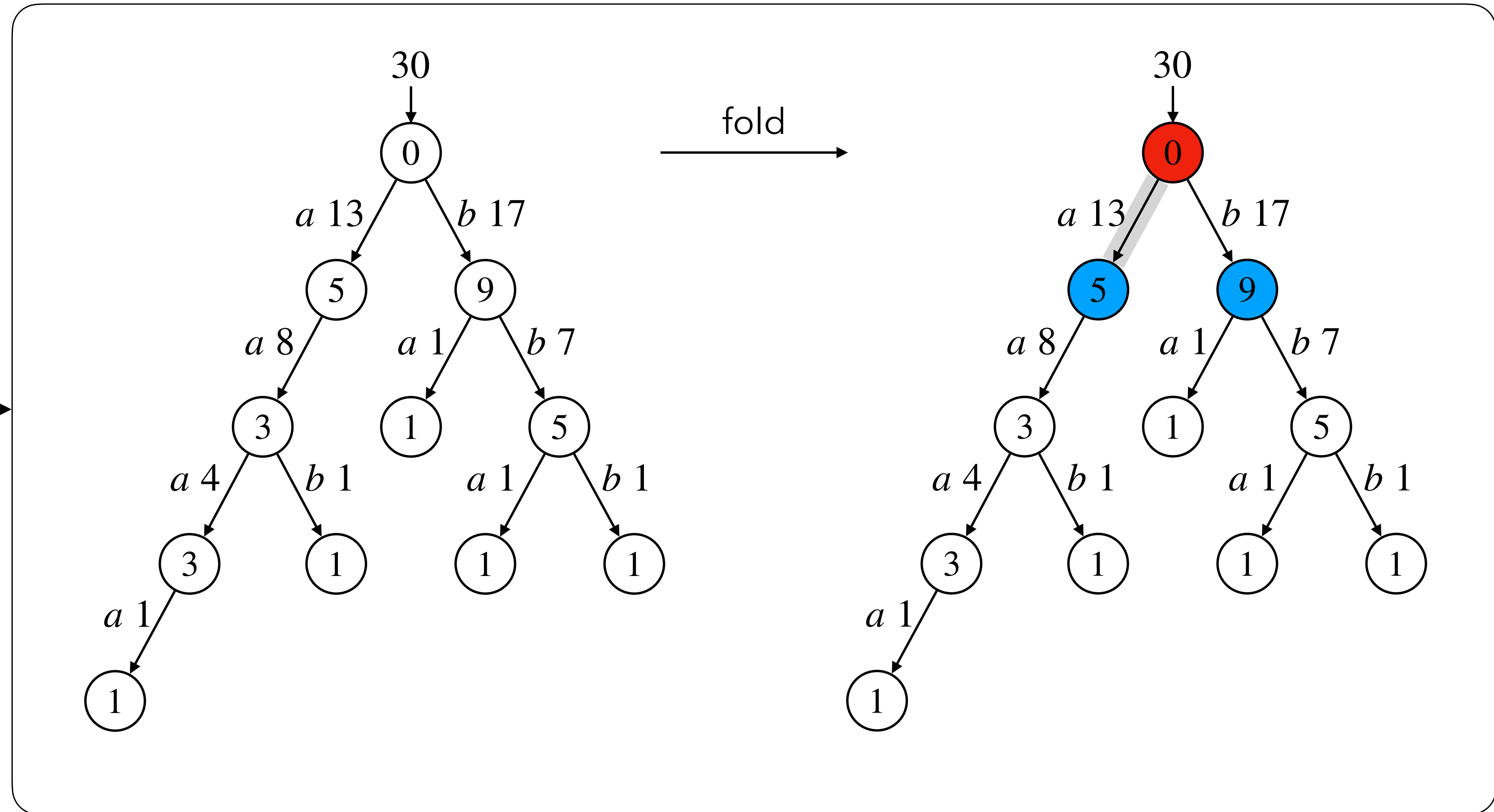
Batch learning of PDFFA: Alergia

do not merge if their probabilities are statistically different

probabilities to stop: $\frac{5}{13} \stackrel{?}{\approx} \frac{0}{30}$

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



- Top-down method

- Merges blue states into red states

- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

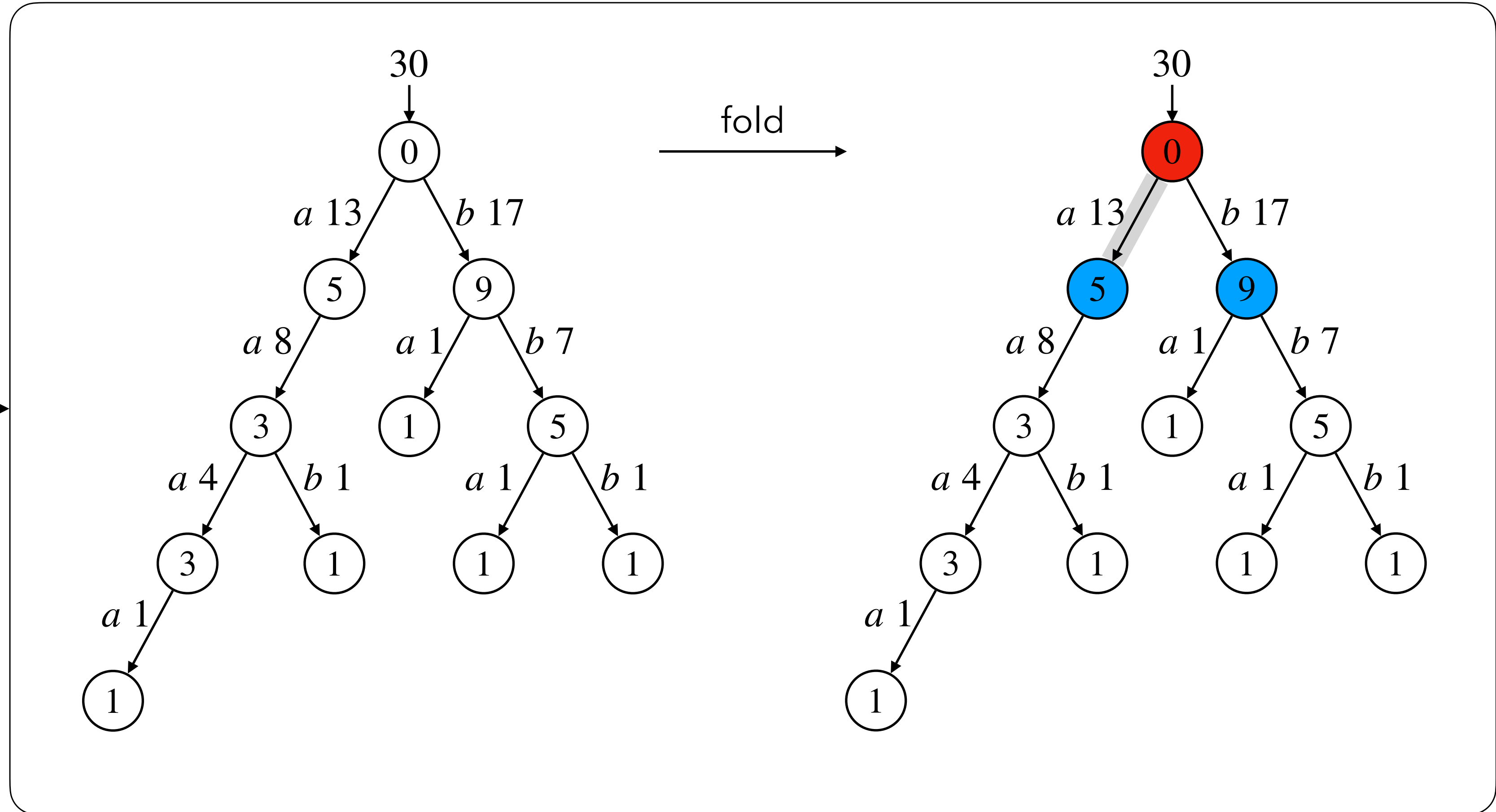
Batch learning of PDFFA: Alergia

do not merge if their probabilities are statistically different

probabilities to stop: $\frac{5}{13} \not\approx \frac{0}{30}$

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



- Top-down method

- Merges blue states into red states

- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

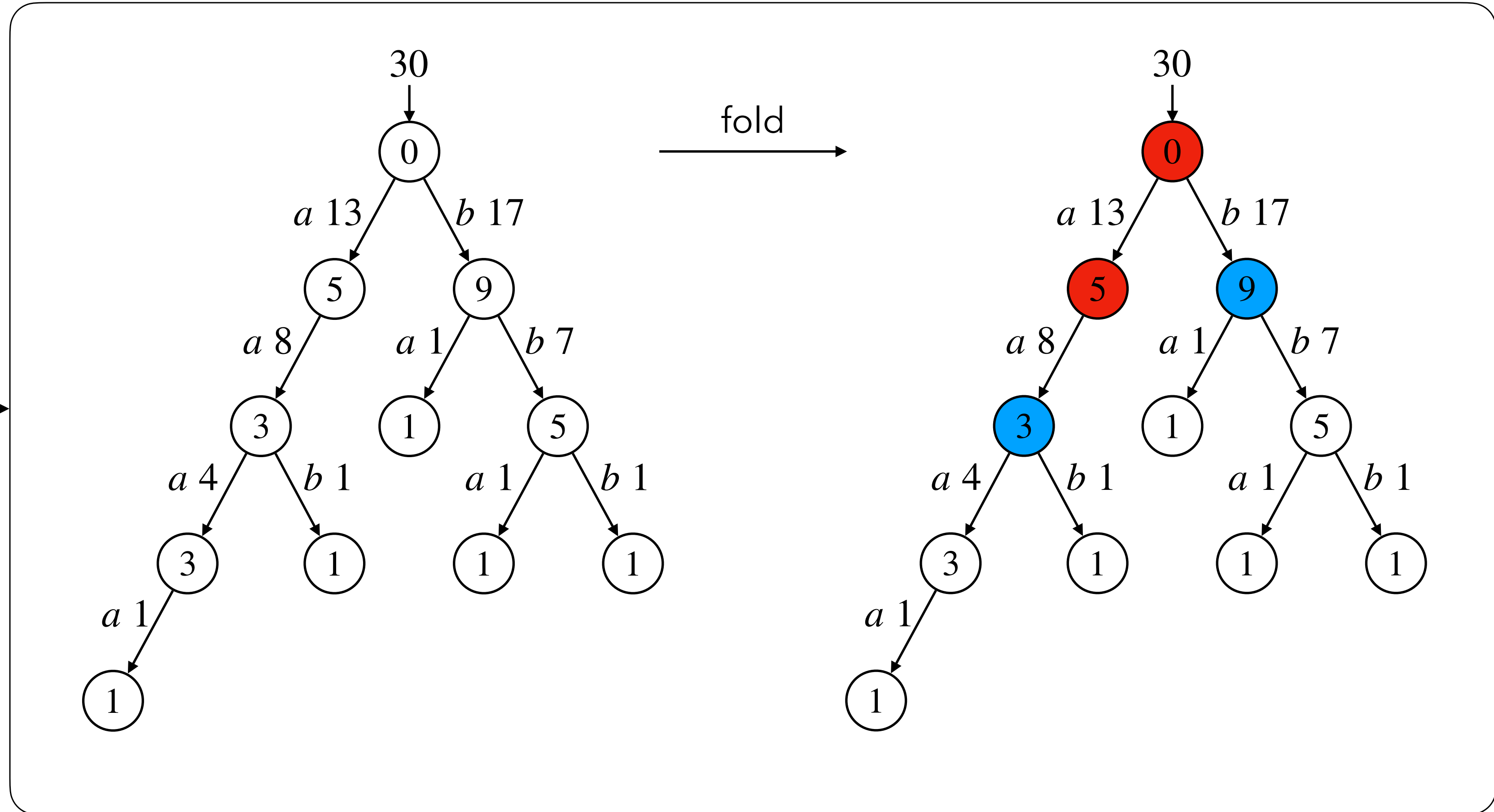
Batch learning of PDFFA: Alergia

do not merge if their probabilities are statistically different

probabilities to stop: $\frac{5}{13} \not\approx \frac{0}{30}$

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



- Top-down method

- Merges **blue states** into **red states**

- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

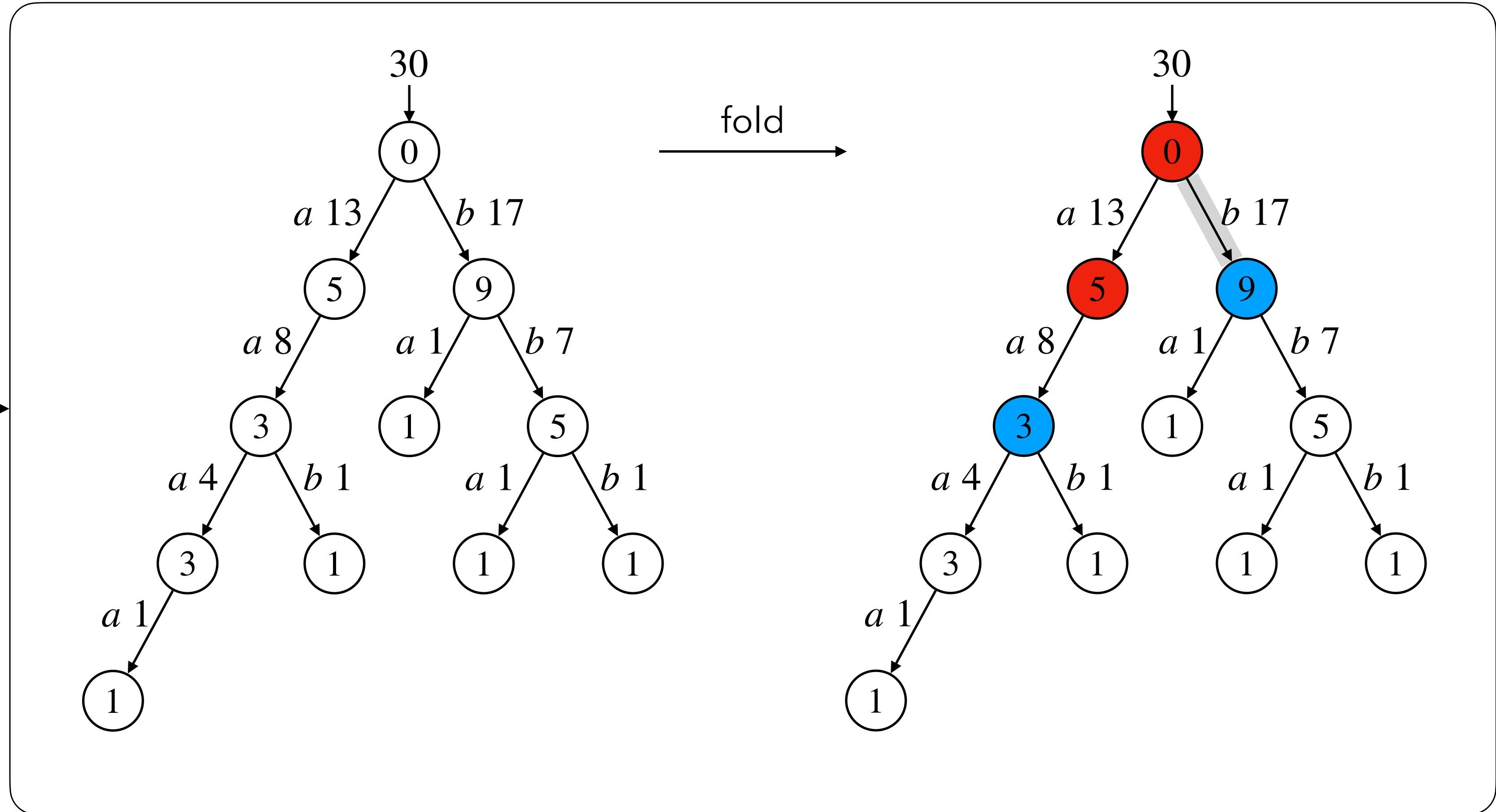
Batch learning of PDFFA: Alergia

do not merge if their probabilities are statistically different

probabilities to stop: $\frac{5}{13} \not\approx \frac{0}{30}$

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



- Top-down method

- Merges blue states into red states

- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

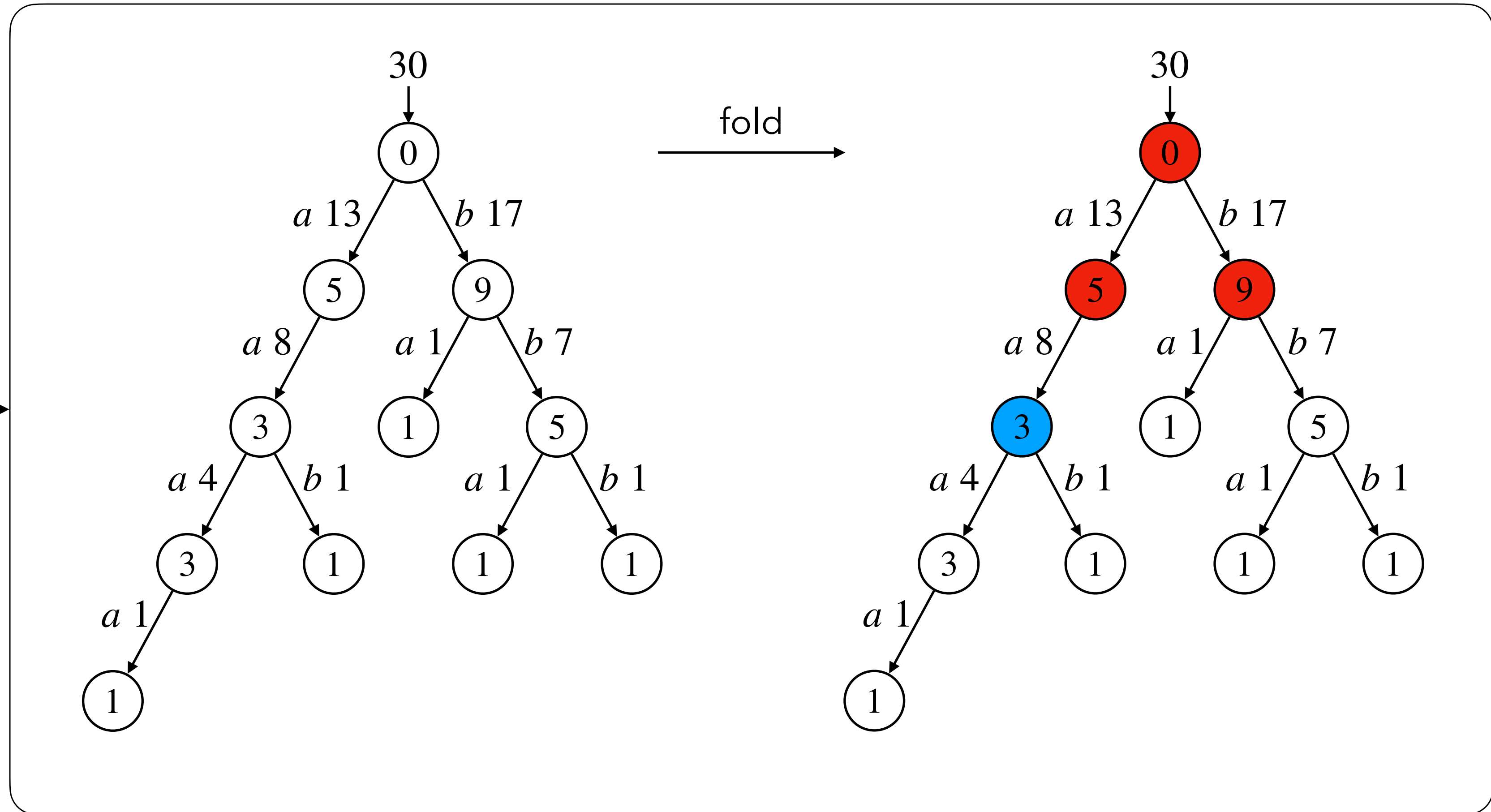
Batch learning of PDFFA: Alergia

do not merge if their probabilities are statistically different

probabilities to stop: $\frac{5}{13} \not\approx \frac{0}{30}$ $\frac{9}{17} \not\approx \frac{0}{30}$

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	$\langle bb \rangle^5$ Leucocytes
I	Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



- Top-down method

- Merges **blue states** into **red states**

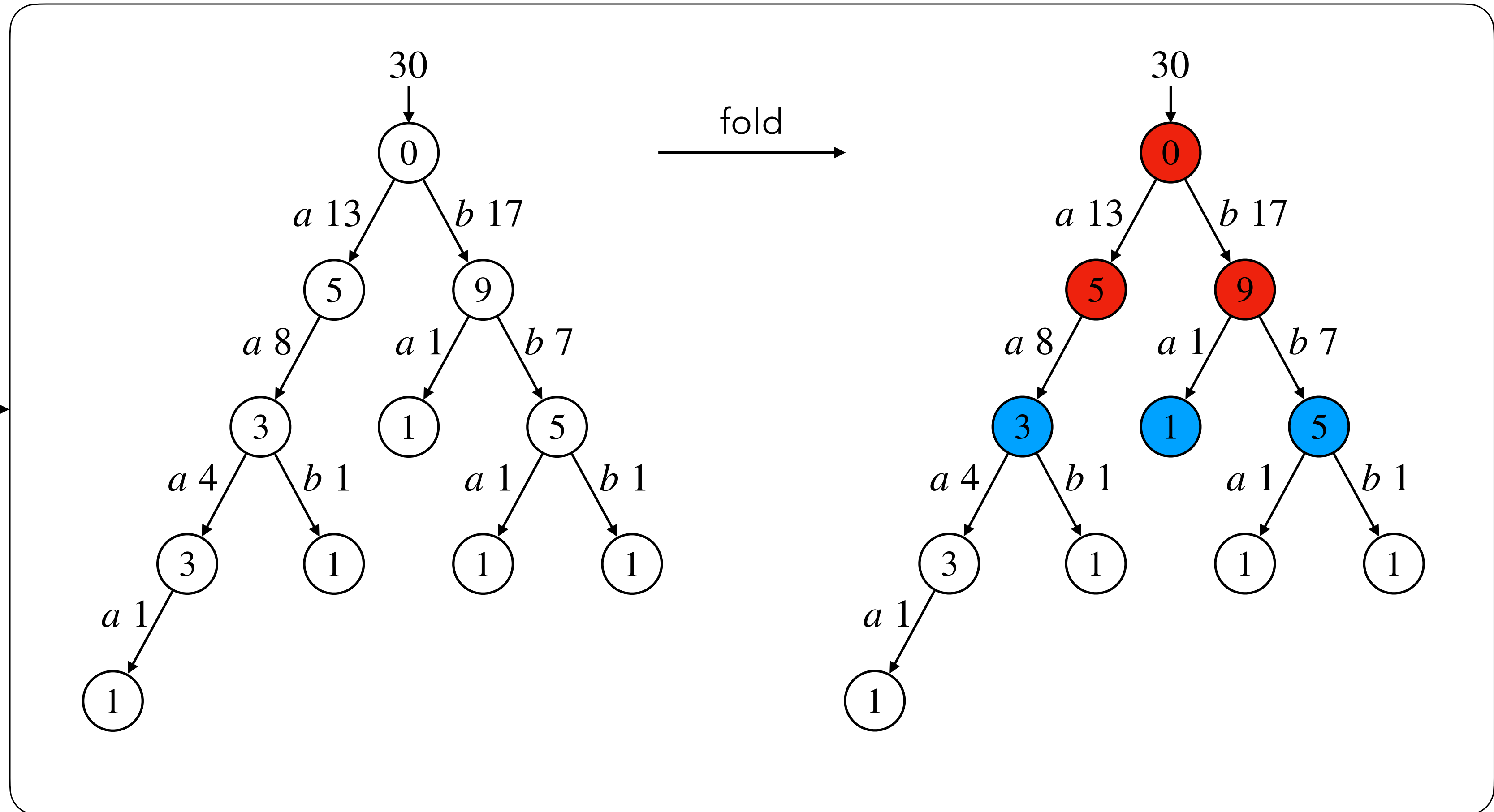
- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



- Top-down method

- Merges blue states into red states

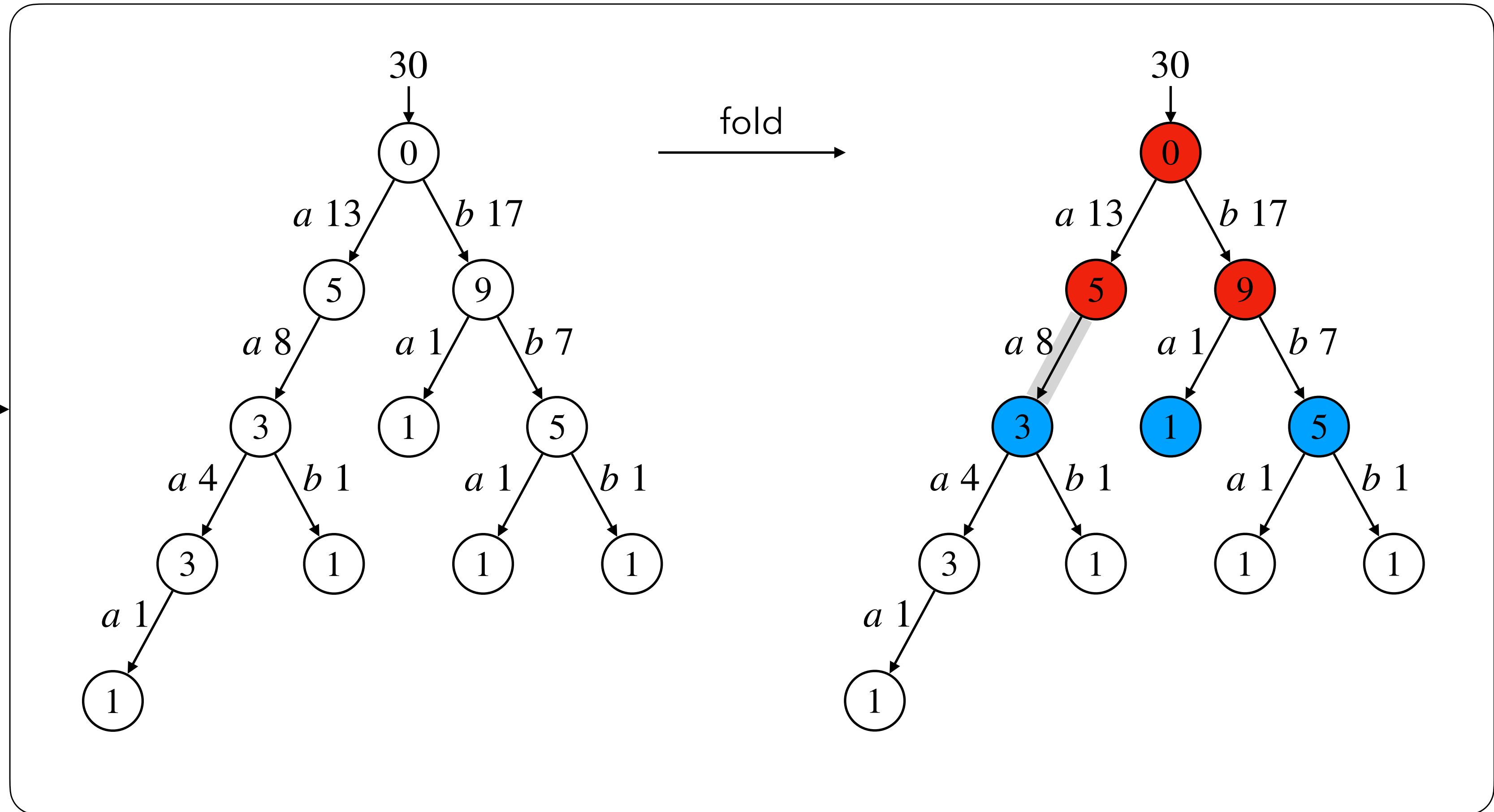
- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



- Top-down method

- Merges **blue states** into **red states**

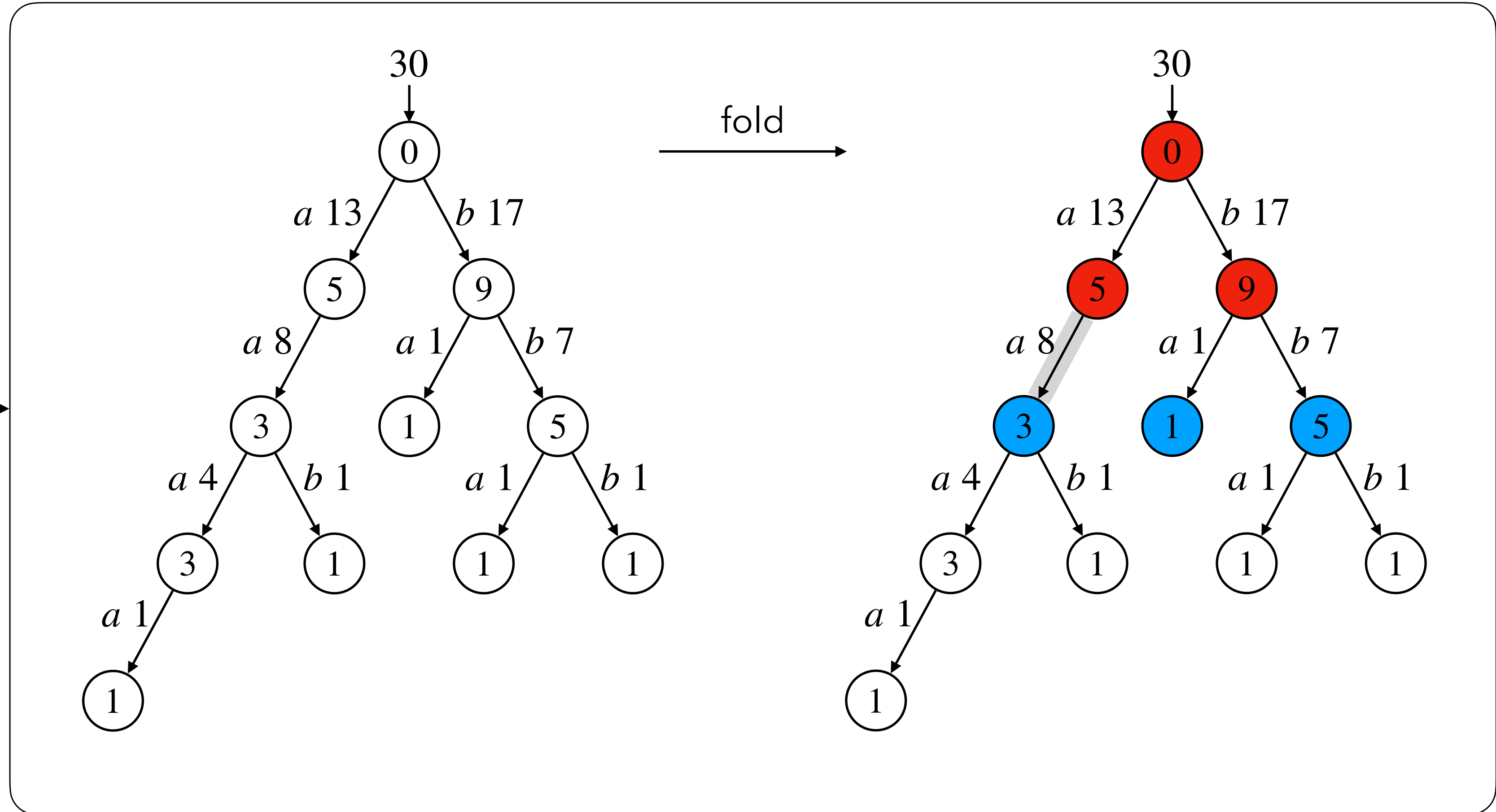
- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	$\langle aaa \rangle^3$ Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	$\langle bb \rangle^5$ Leucocytes
I	Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	$\langle bbb \rangle^1$ CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



stop $\frac{3}{8} \approx \frac{5}{13}$ a $\frac{4}{8} \approx \frac{8}{13}$ b $\frac{1}{8} \approx \frac{0}{13}$

- Top-down method

- Merges blue states into red states

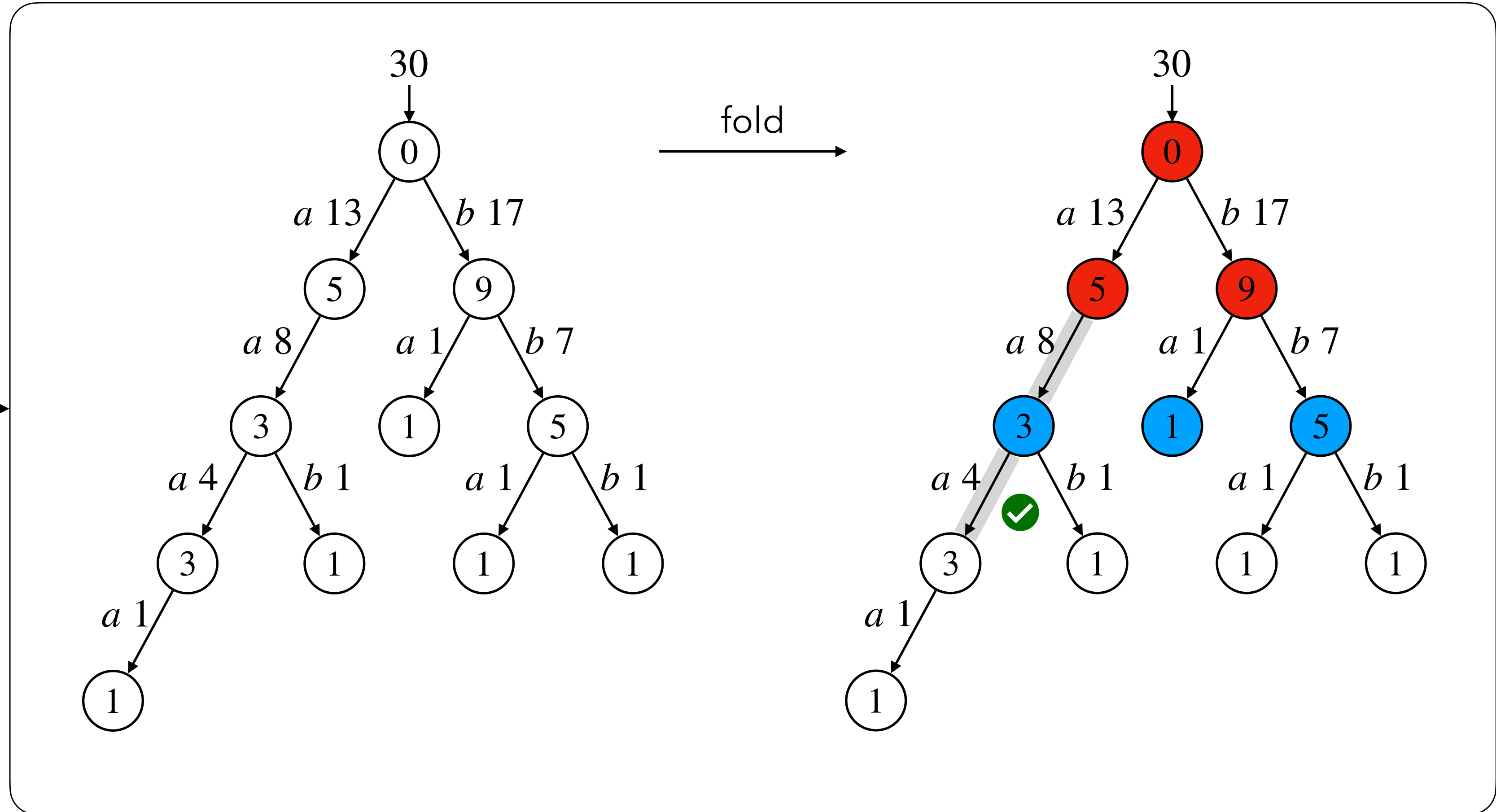
- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	$\langle aaa \rangle^3$ Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	$\langle bb \rangle^5$ Leucocytes
I	Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	$\langle bbb \rangle^1$ CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



$$\begin{array}{ccc} \text{stop} & a & b \\ \frac{3}{8} \approx \frac{5}{13} & \frac{4}{8} \approx \frac{8}{13} & \frac{1}{8} \approx \frac{0}{13} \end{array}$$

- Top-down method

- Merges blue states into red states

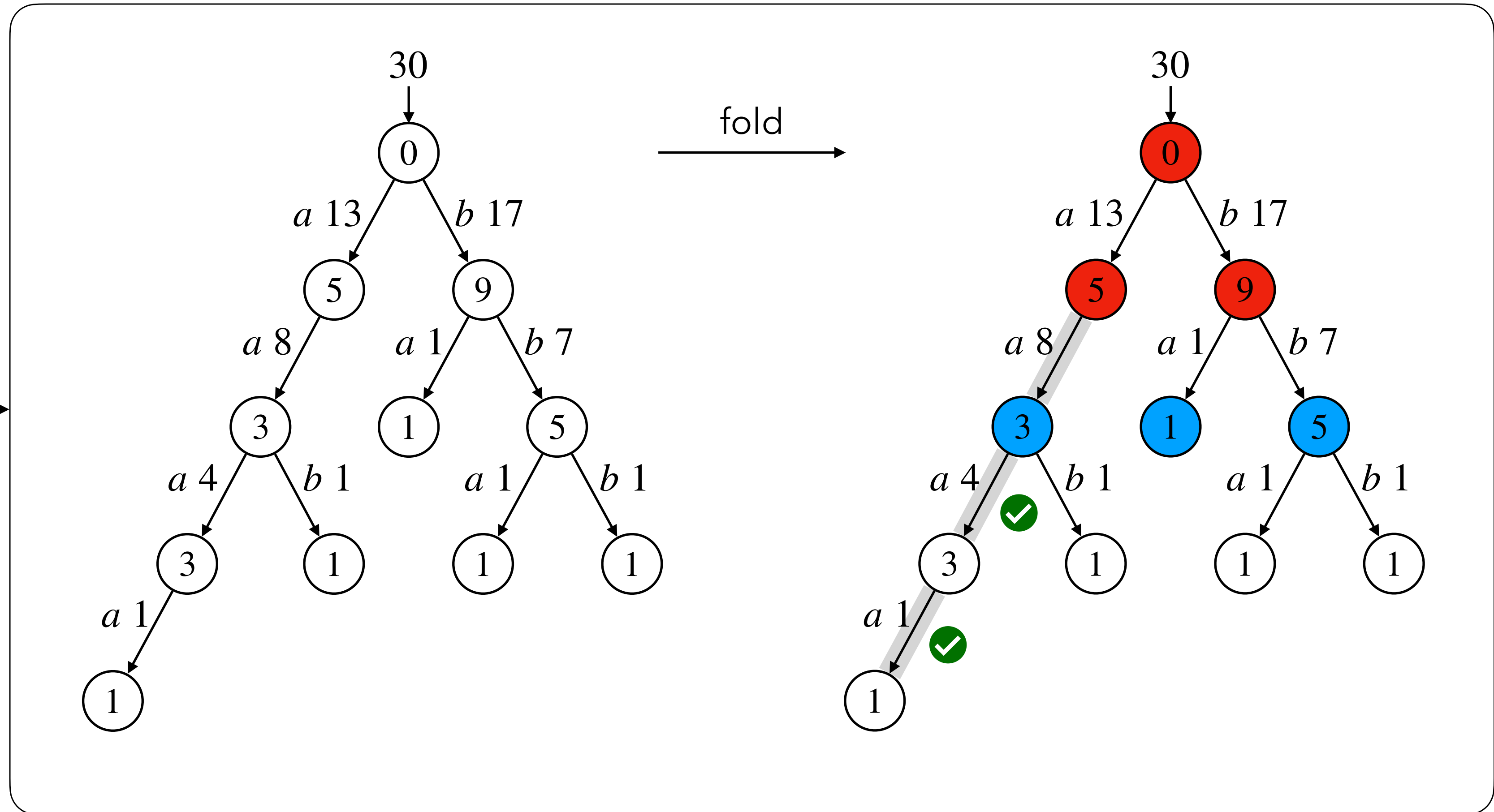
- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	$\langle aaa \rangle^3$ Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	$\langle bb \rangle^5$ Leucocytes
I	Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	$\langle bbb \rangle^1$ CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



$$\begin{array}{ccc} \text{stop} & a & b \\ \frac{3}{8} \approx \frac{5}{13} & \frac{4}{8} \approx \frac{8}{13} & \frac{1}{8} \approx \frac{0}{13} \end{array}$$

- Top-down method

- Merges blue states into red states

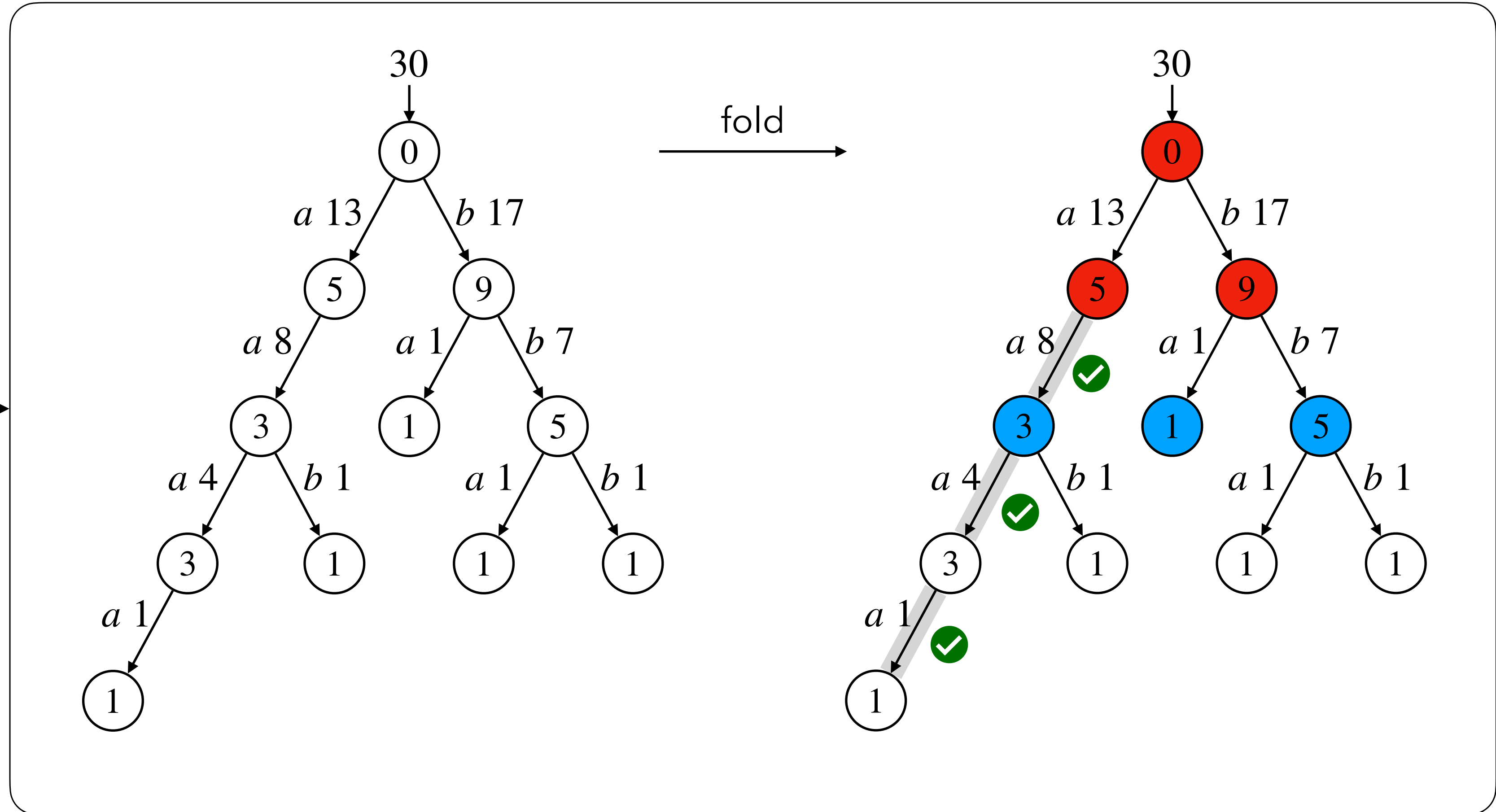
- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	$\langle aaa \rangle^3$ Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



stop $\frac{3}{8} \approx \frac{5}{13}$ a $\frac{4}{8} \approx \frac{8}{13}$ b $\frac{1}{8} \approx \frac{0}{13}$

- Top-down method
- Merges **blue states** into **red states**

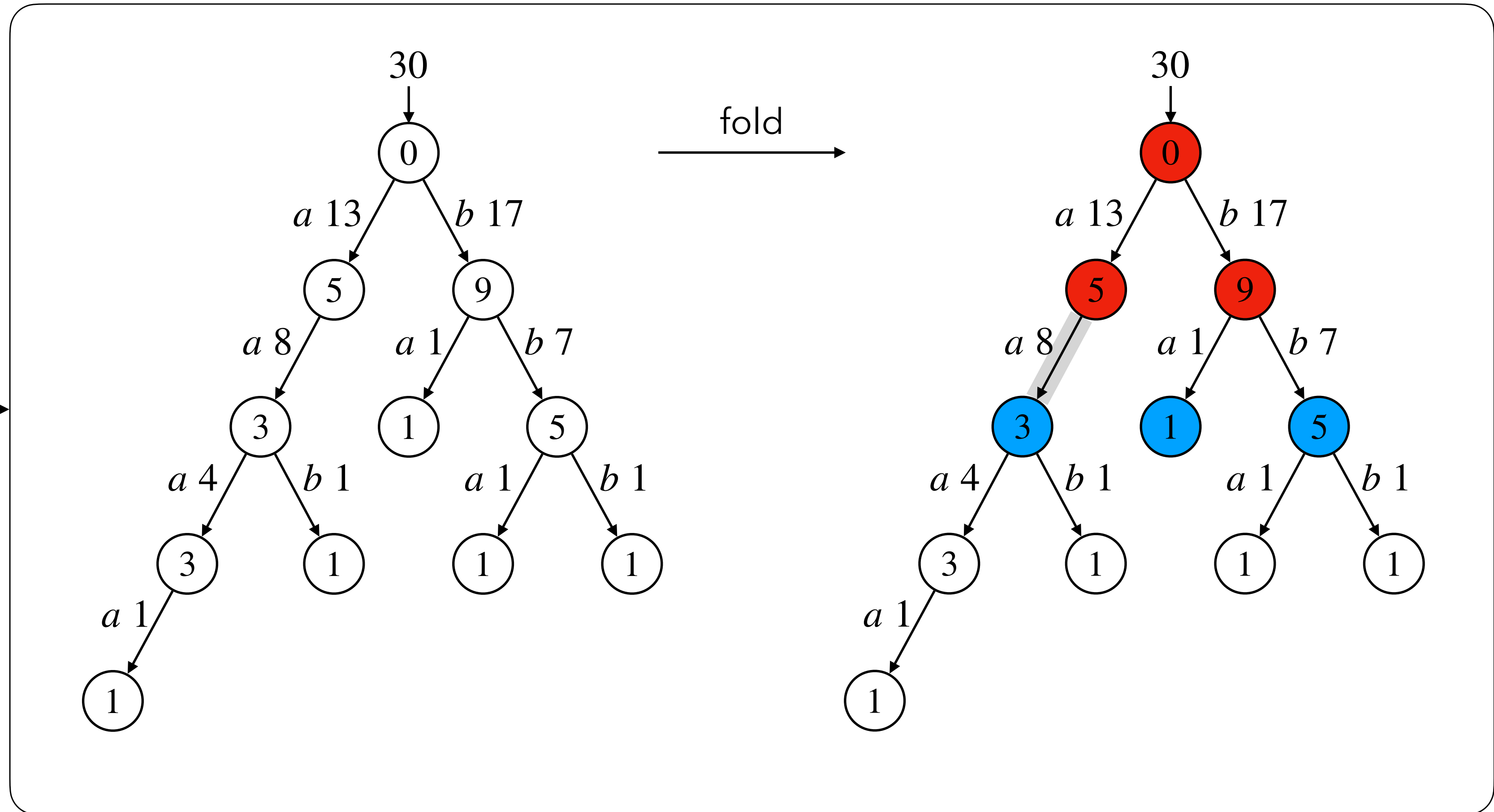
- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	$\langle aaa \rangle^3$ Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	$\langle bb \rangle^5$ Leucocytes
I	Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	$\langle bbb \rangle^1$ Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



- Top-down method

- Merges blue states into red states

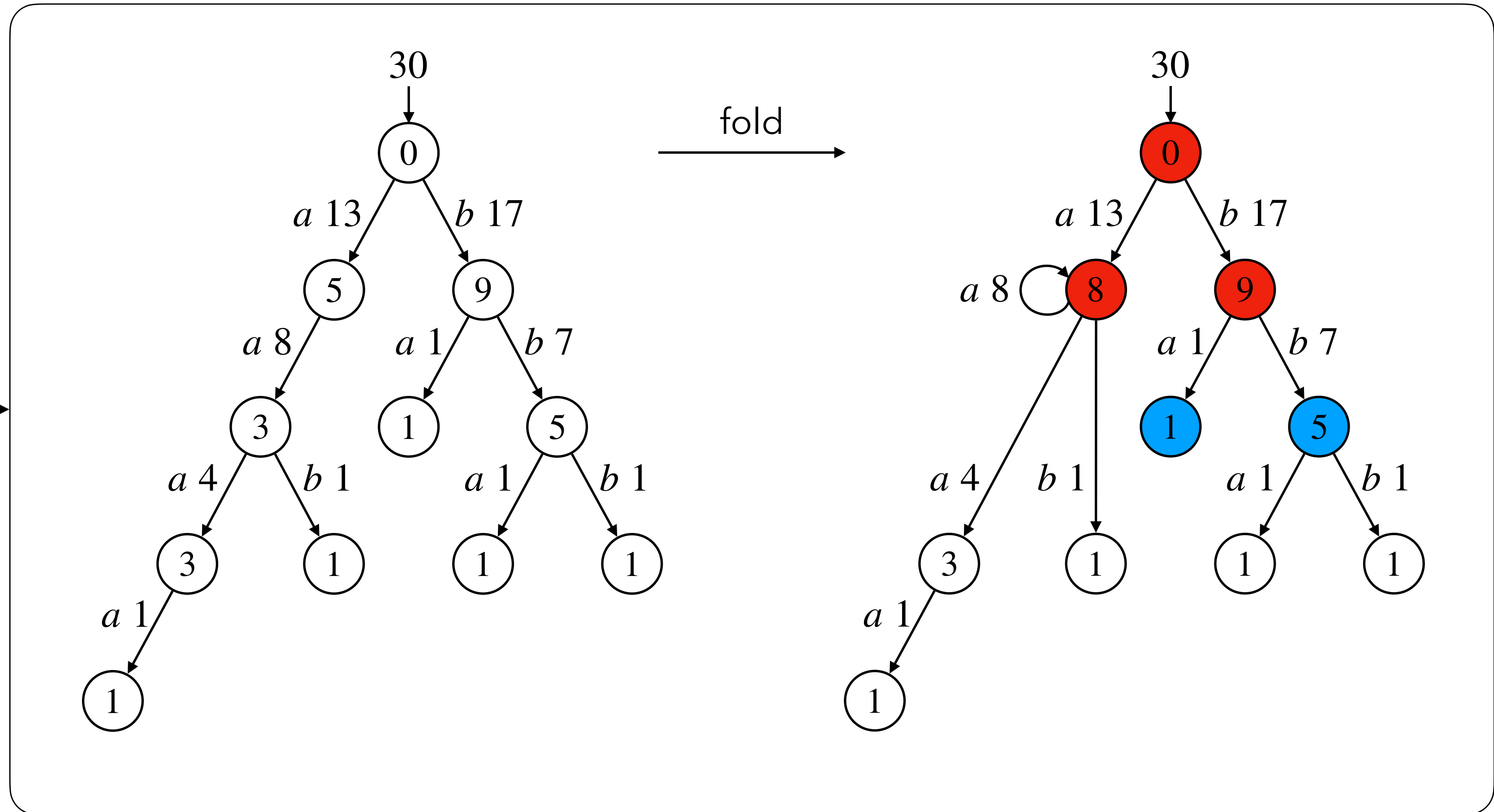
- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	$\langle aaa \rangle^3$ Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	$\langle bb \rangle^5$ Leucocytes
I	Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	$\langle bbb \rangle^1$ CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



- Top-down method

- Merges blue states into red states

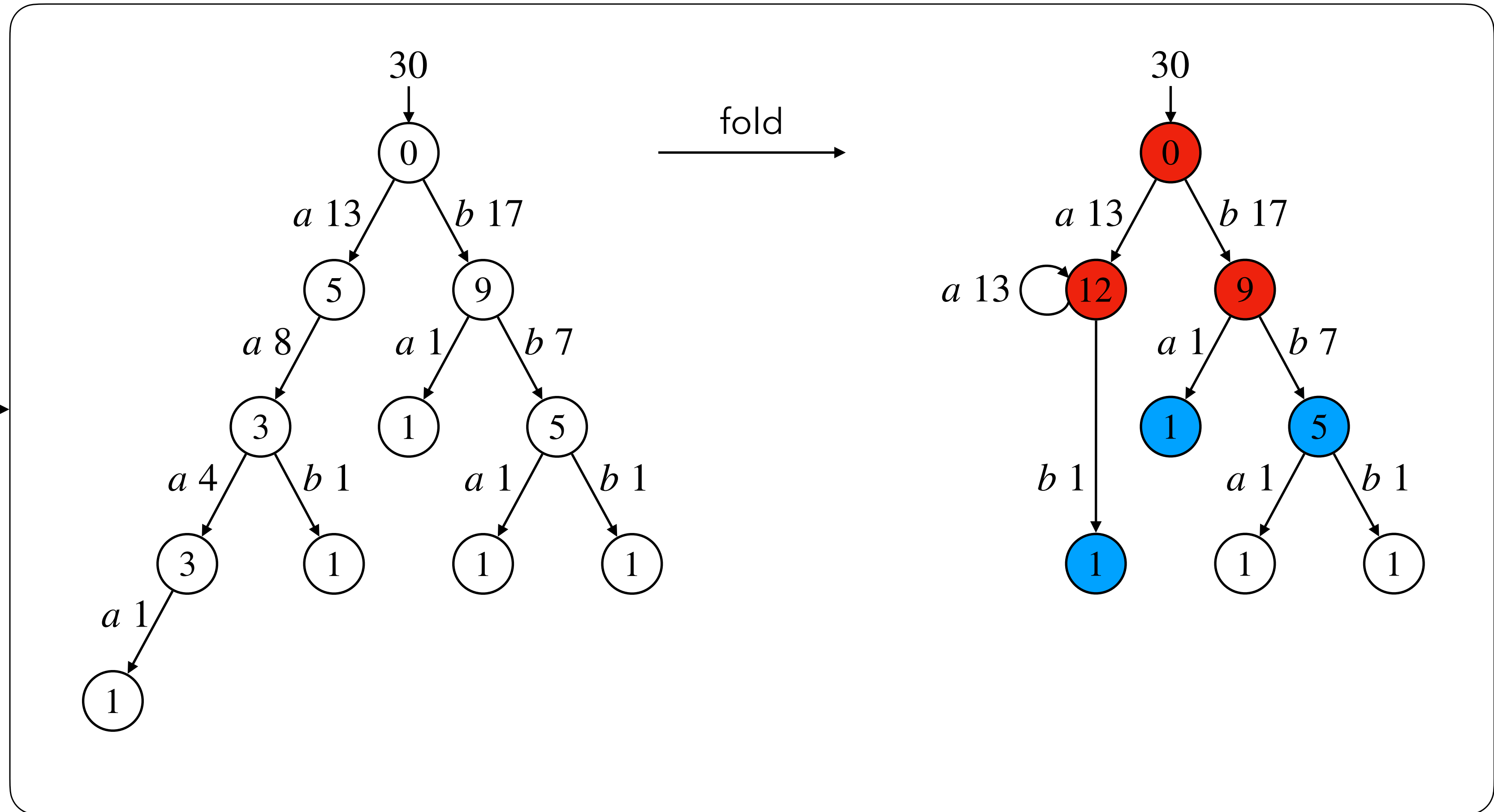
- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

Batch learning of PDFFA: Alergia

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	$\langle aaa \rangle^3$ Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	$\langle bb \rangle^5$ Leucocytes
I	Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	$\langle bbb \rangle^1$ CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration

train



- Top-down method

- Merges blue states into red states

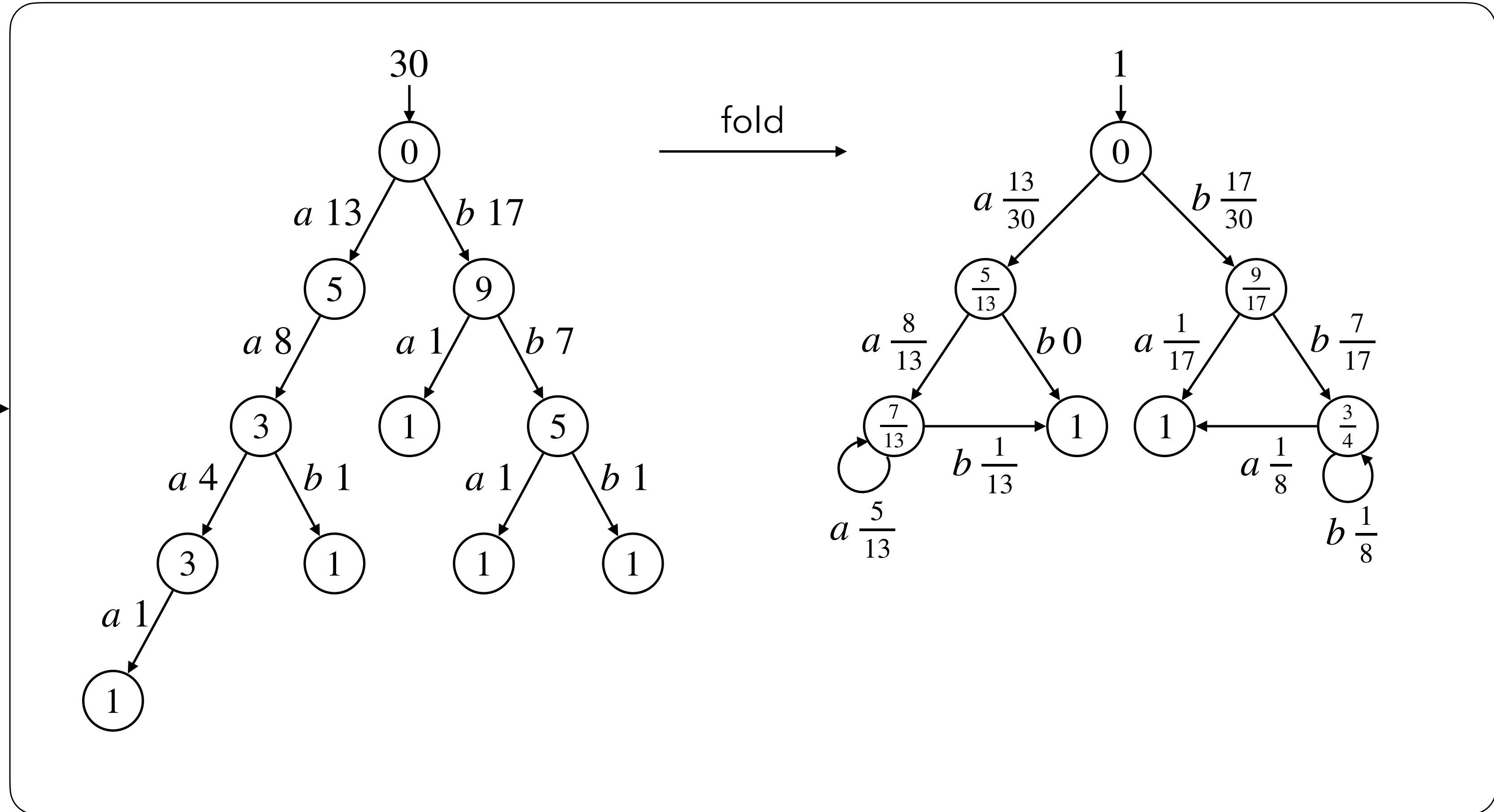
- Hoeffding incompatibility

$$\left| \frac{x_1}{n_1} - \frac{x_2}{n_2} \right| \geq \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} \left(\frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_2}} \right)$$

Batch learning of PDFFA: n-gram

case:concept:name	concept:name
XJ	$\langle a \rangle^5$ ER Registration
XJ	ER Triage
XJ	$\langle aa \rangle^3$ ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	$\langle aaa \rangle^3$ CRP
XJ	IV Liquid
XJ	$\langle aab \rangle^1$ IV Antibiotics
XJ	Admission NC
XJ	$\langle aaaa \rangle^1$ Leucocytes
I	ER Registration
I	$\langle b \rangle^9$ ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	$\langle ba \rangle^1$ LacticAcid
I	CRP
I	Leucocytes
I	$\langle bb \rangle^5$ Admission NC
WEA	ER Registration
WEA	$\langle bba \rangle^1$ ER Triage
WEA	ER Sepsis Triage
WEA	$\langle bbb \rangle^1$ CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



Streaming Learning

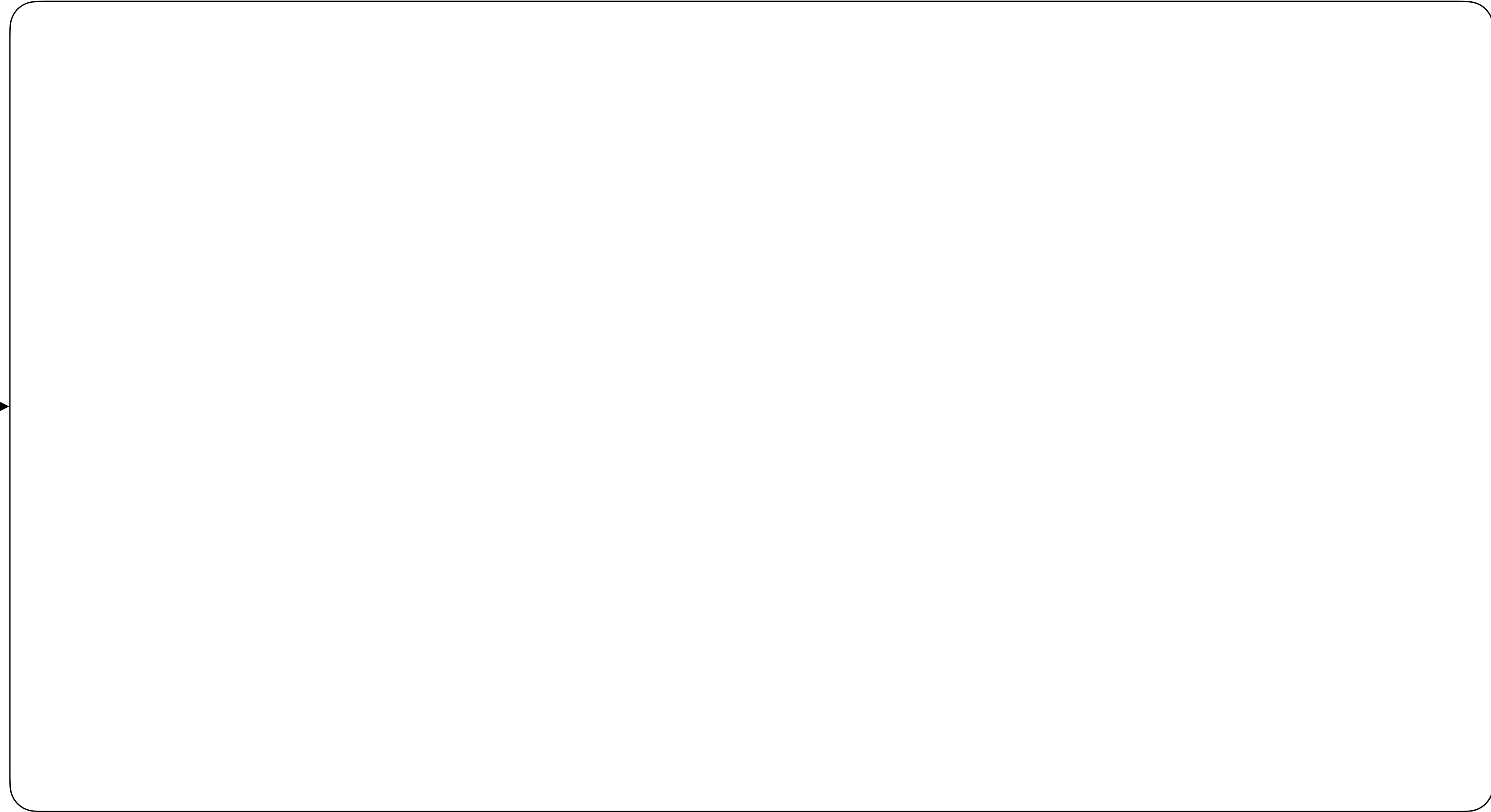
Batch learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

event-log

$$L : \Sigma^* \rightarrow \mathcal{N}$$

train



Batch learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

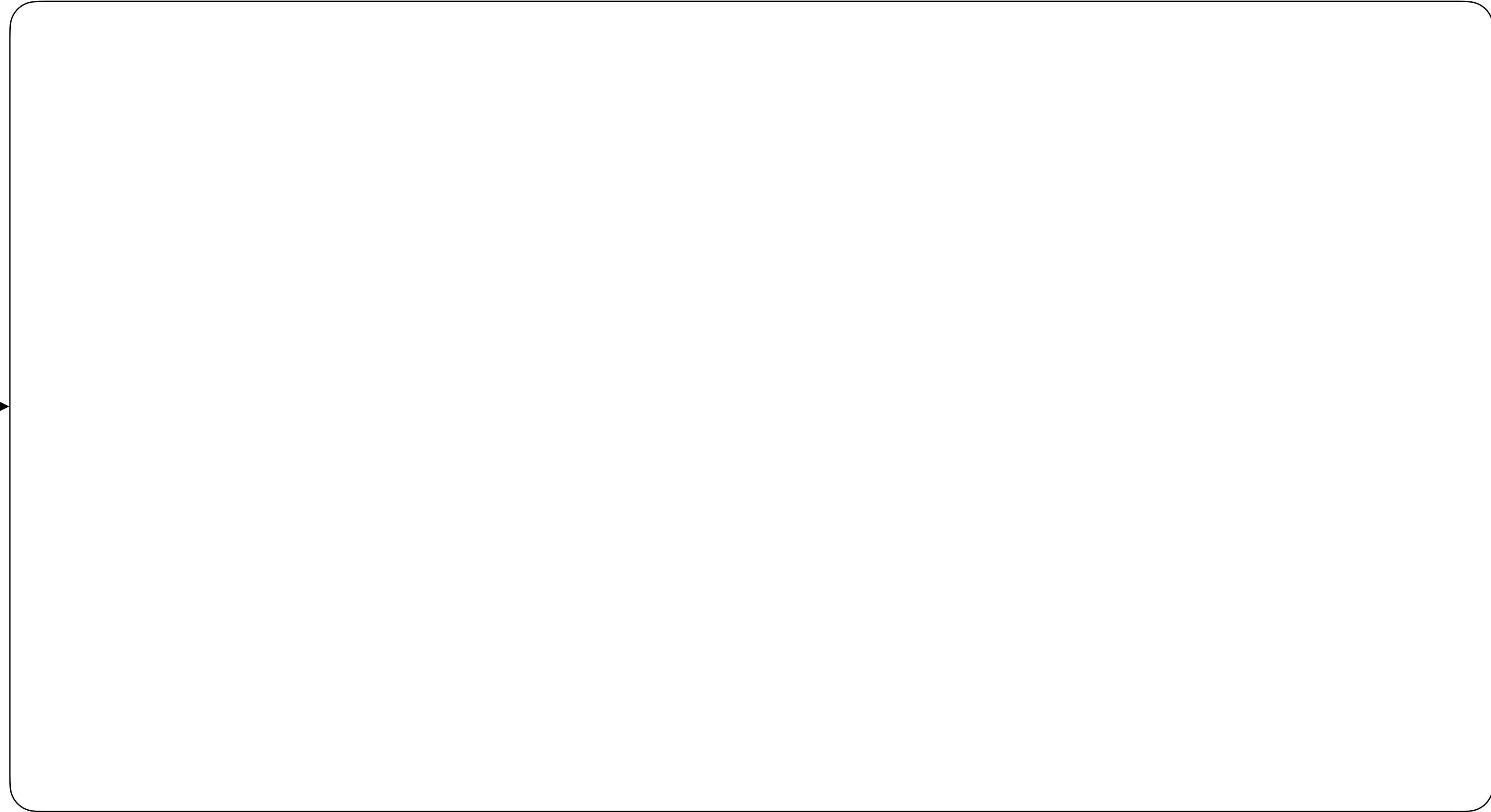
train



Streaming learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



Streaming learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train

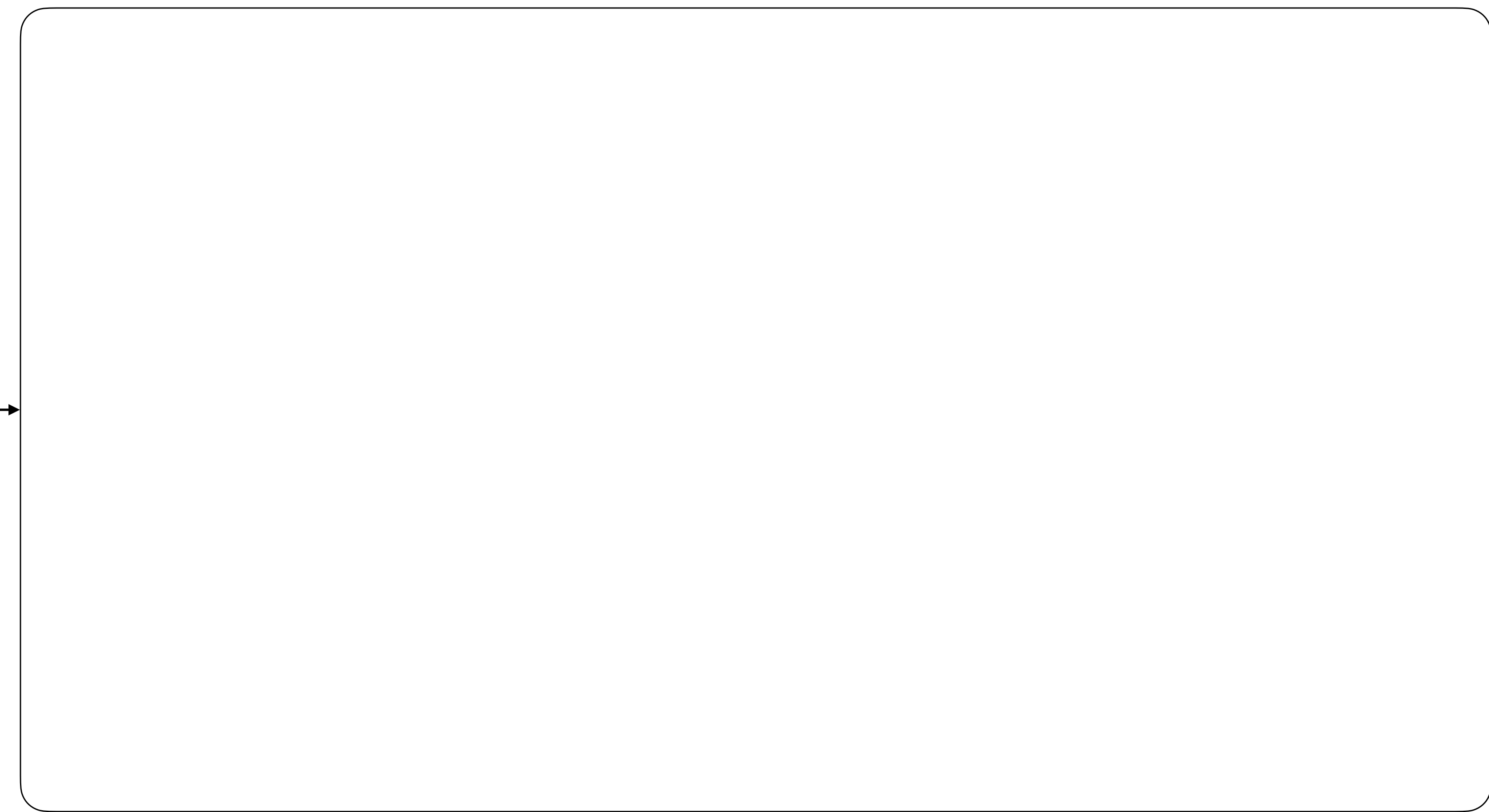


- Events are given one by one, interleaved across cases.

Streaming learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train



- Events are given one by one, interleaved across cases.
- Training is conducted online and within the running system.

Streaming learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train

maintain current state for every active case id

$$p: \begin{cases} Q \rightarrow \Delta(\Sigma \cup \{\text{stop}\}) \\ Q \times \Sigma \rightarrow Q \end{cases} \quad \text{incremental}$$

- Events are given one by one, interleaved across cases
- Training is conducted online and within the running system.

Streaming learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train

case id

inference

maintain current state for every active case id

$$p: \begin{cases} Q \rightarrow \Delta(\Sigma \cup \{\text{stop}\}) \\ Q \times \Sigma \rightarrow Q \end{cases}$$

incremental

?

Streaming learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train

case id

inference

fast & memory-efficient

maintain current state for every active case id

$$p: \begin{cases} Q \rightarrow \Delta(\Sigma \cup \{\text{stop}\}) \\ Q \times \Sigma \rightarrow Q \end{cases}$$

incremental

?

Streaming learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

train

case id

inference

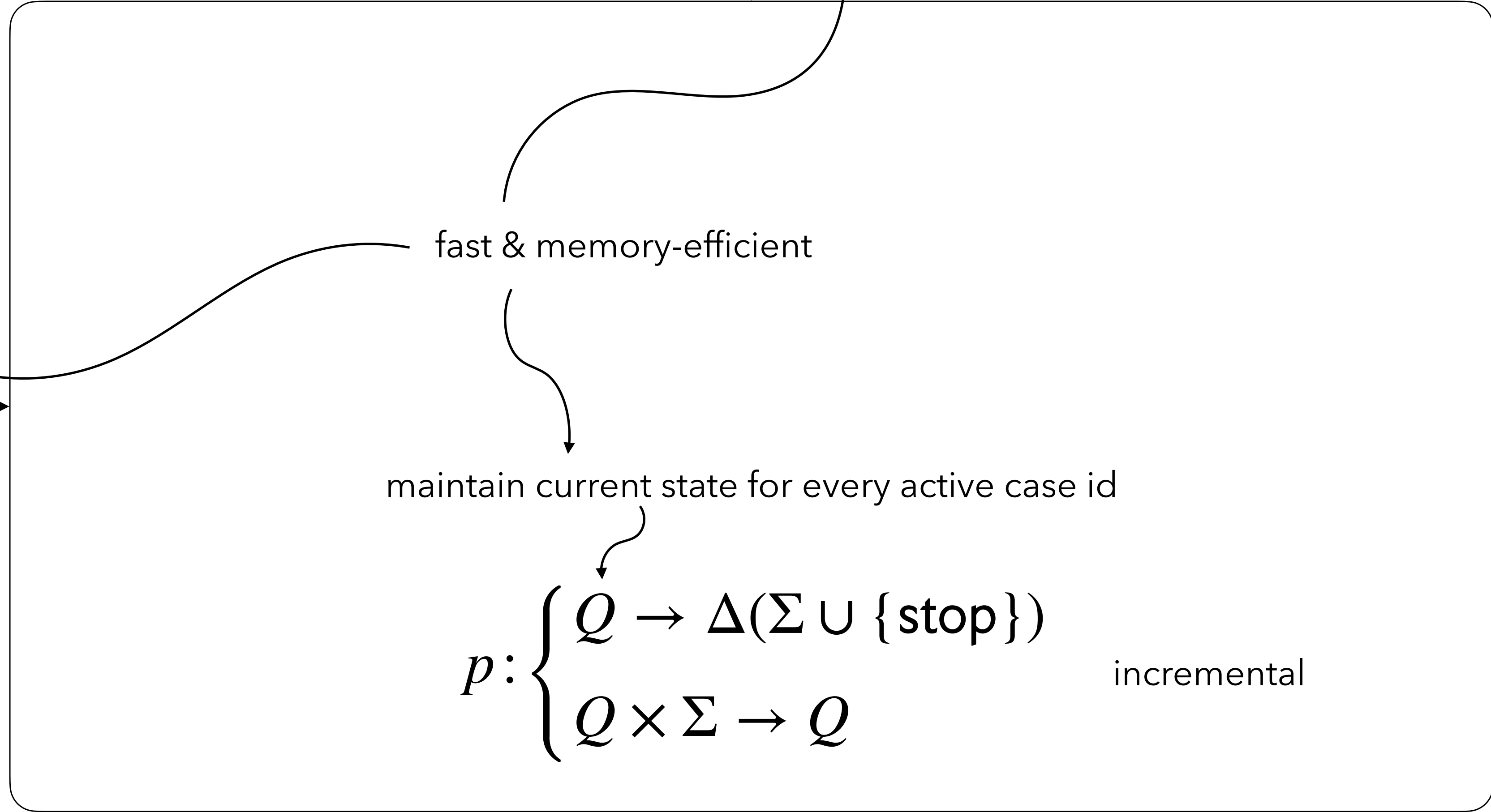
fast & memory-efficient

maintain current state for every active case id

$$p: \begin{cases} Q \rightarrow \Delta(\Sigma \cup \{\text{stop}\}) \\ Q \times \Sigma \rightarrow Q \end{cases}$$

incremental

?



Streaming learning

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

case id

inference

fast & memory-efficient

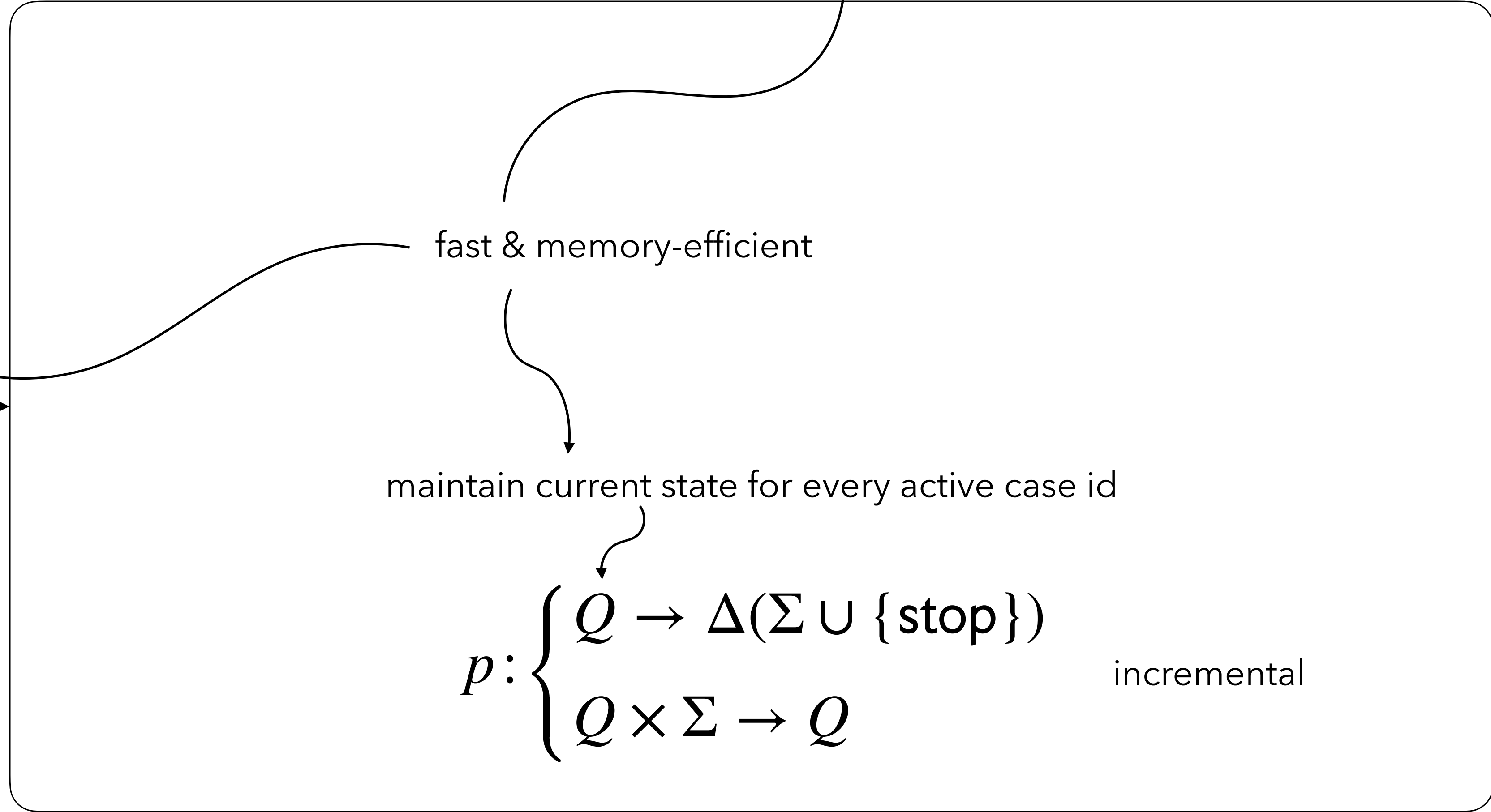
maintain current state for every active case id

$$p: \begin{cases} Q \rightarrow \Delta(\Sigma \cup \{\text{stop}\}) \\ Q \times \Sigma \rightarrow Q \end{cases}$$

incremental

update

?



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

update

3-gram

Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

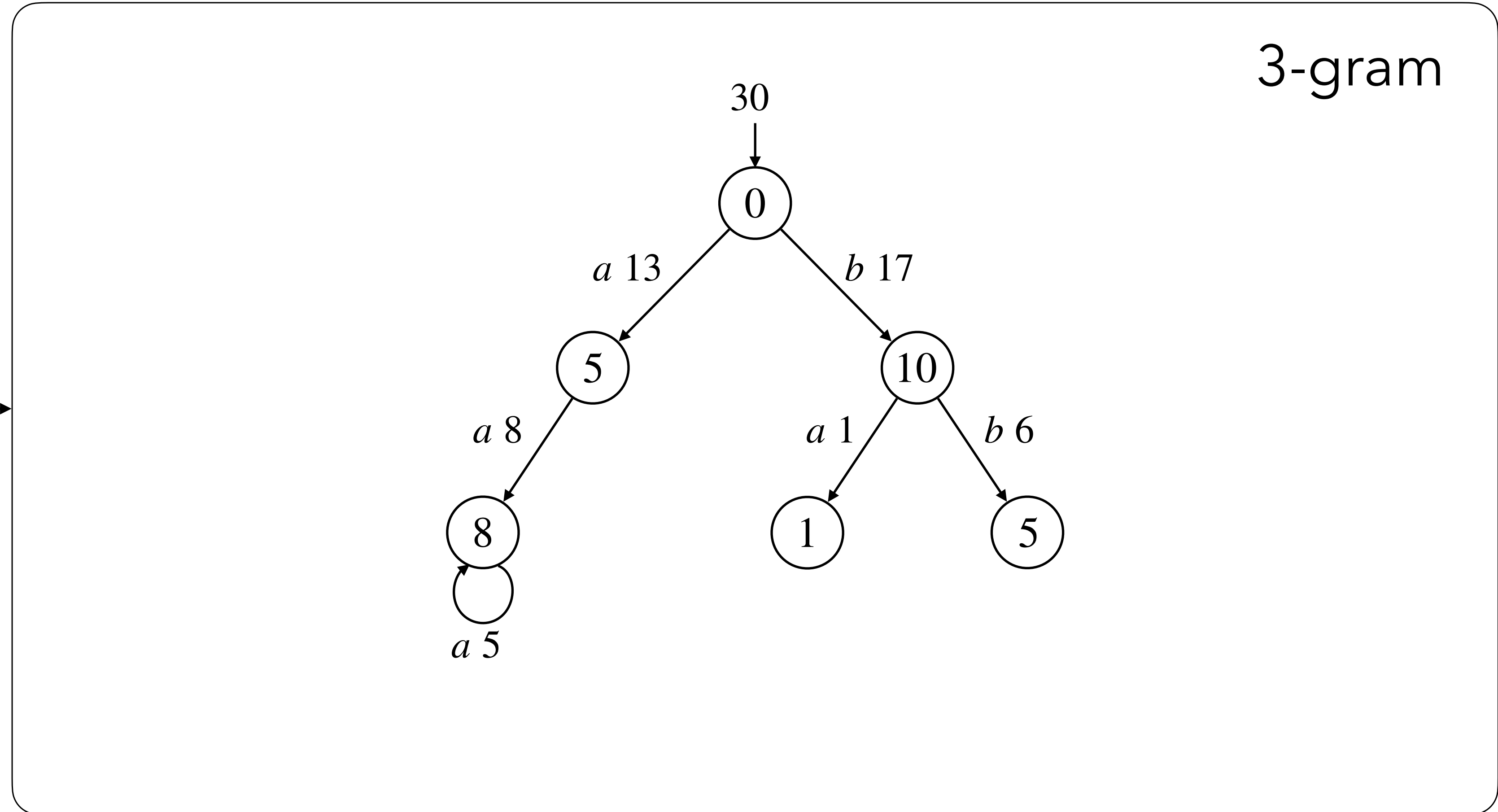
$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

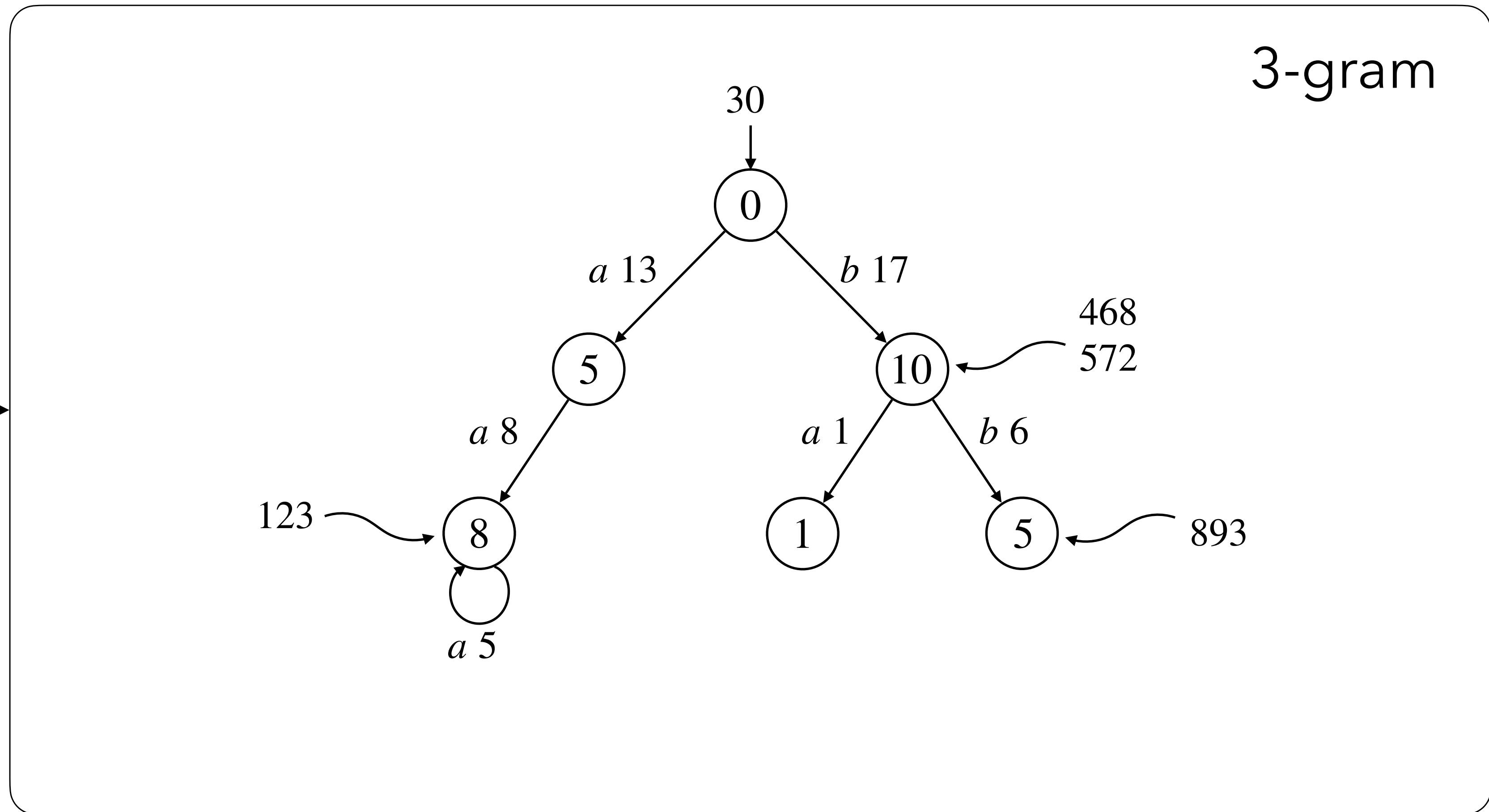
$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

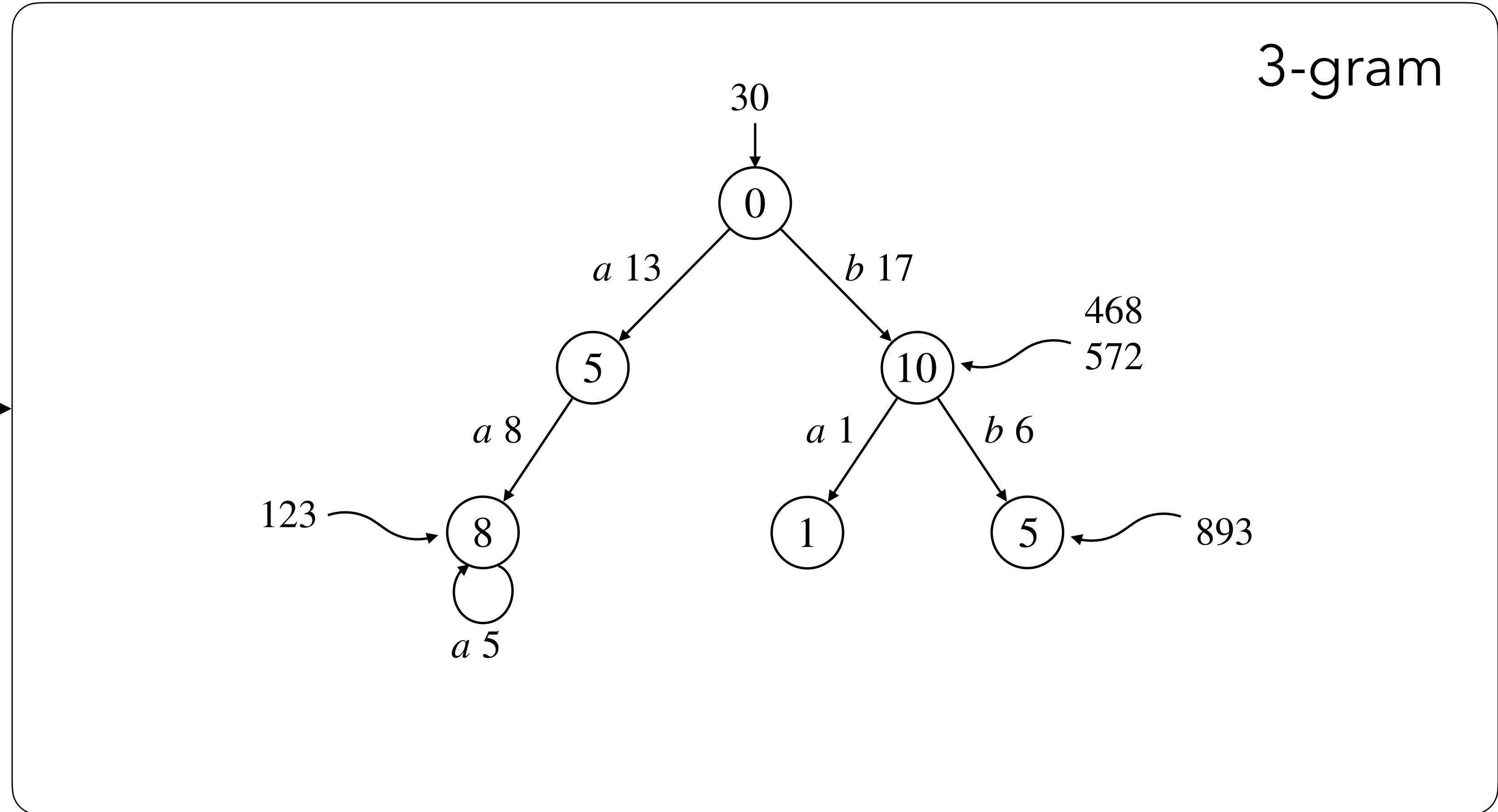
$\langle 123 : a \rangle$

$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

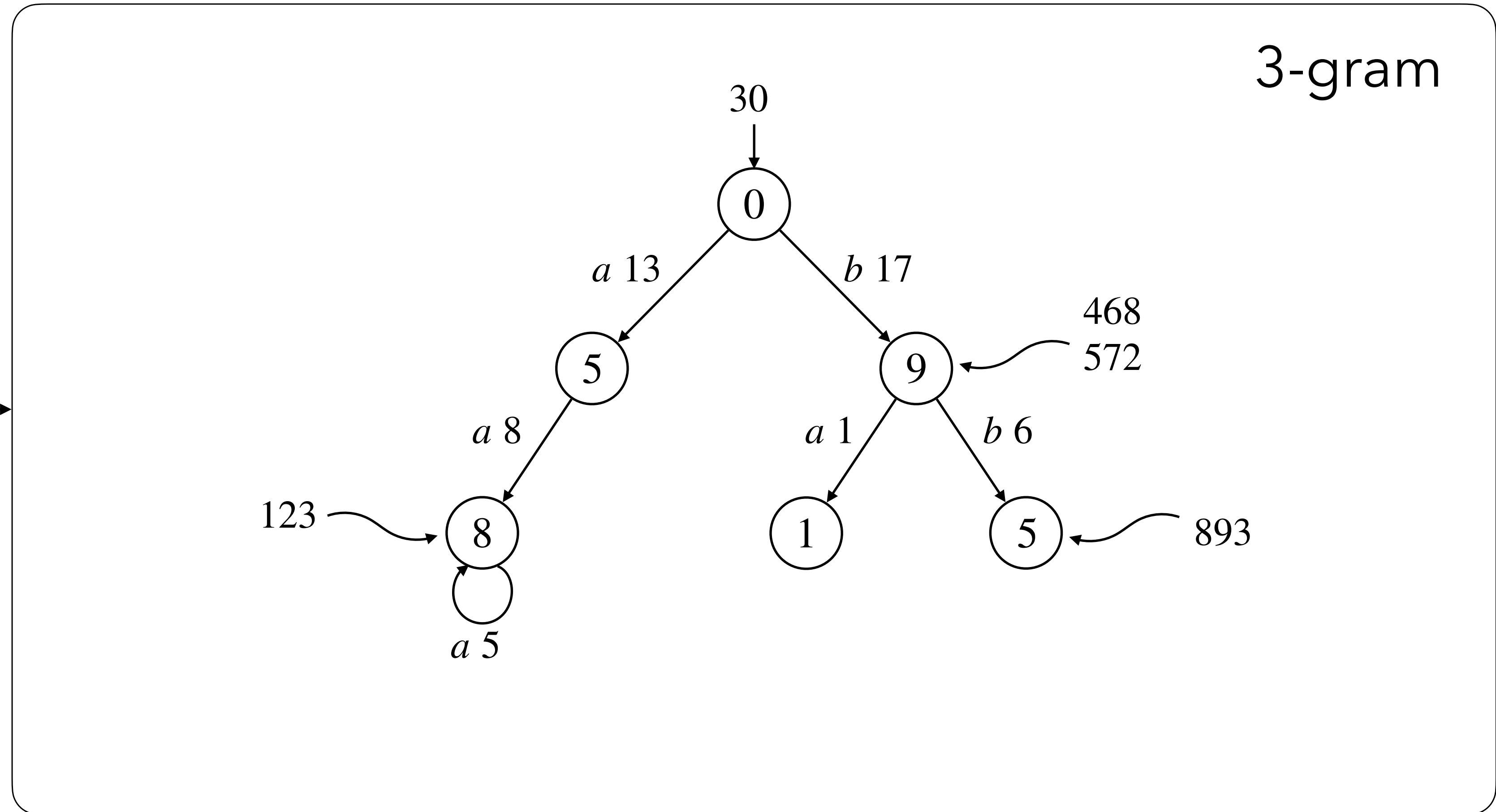
$\langle 123 : a \rangle$

$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

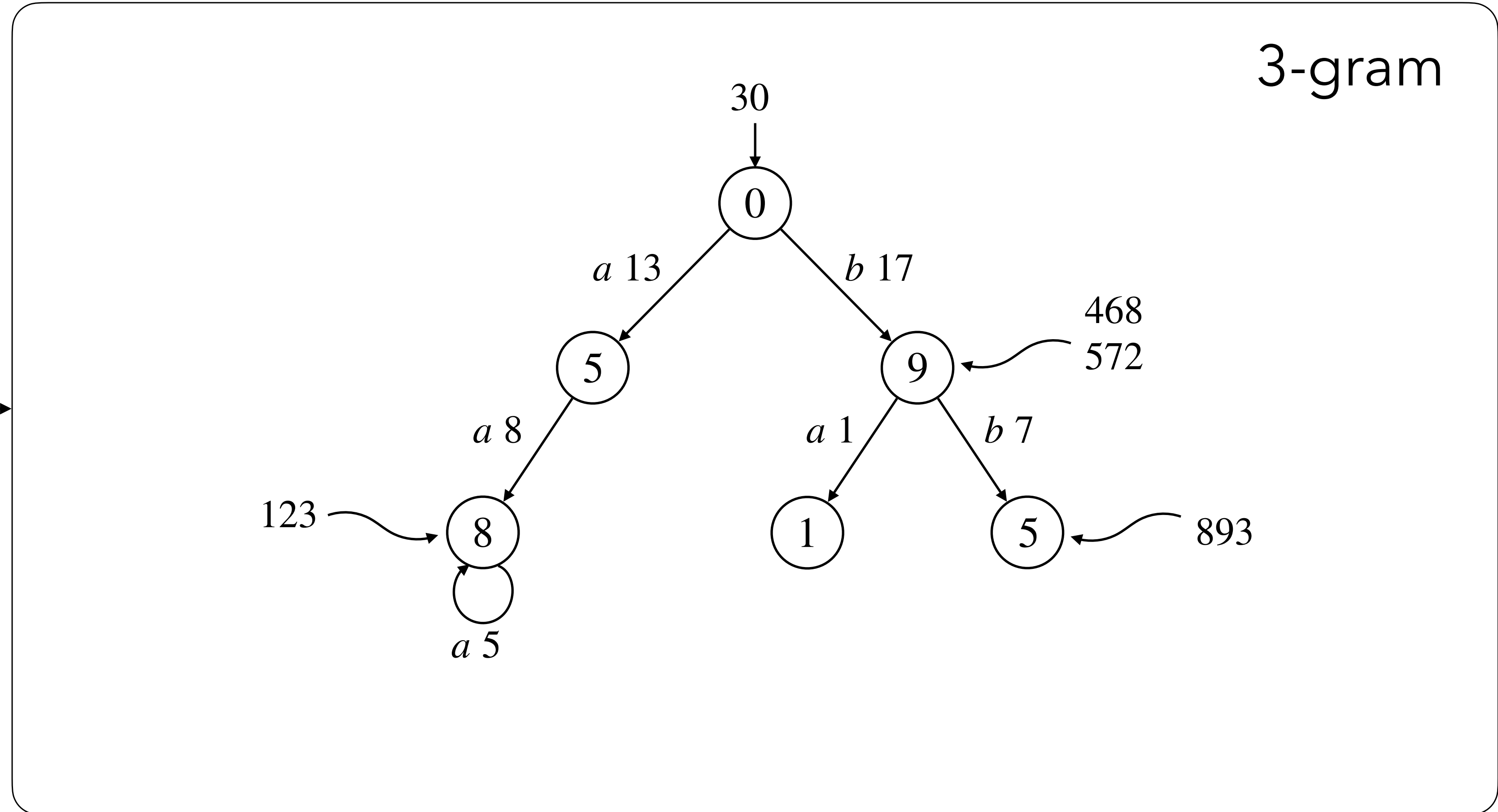
$\langle 123 : a \rangle$

$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

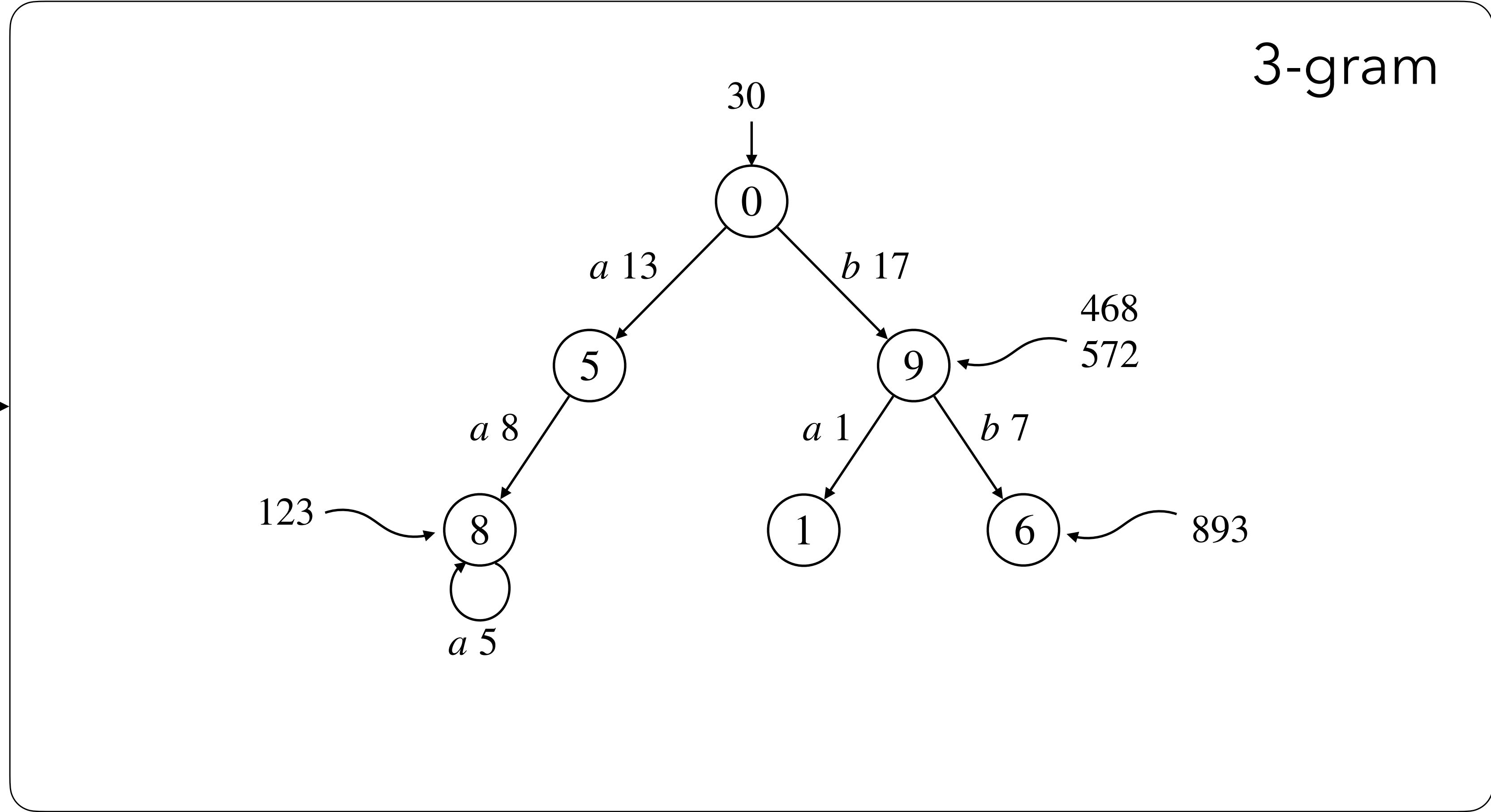
$\langle 123 : a \rangle$

$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

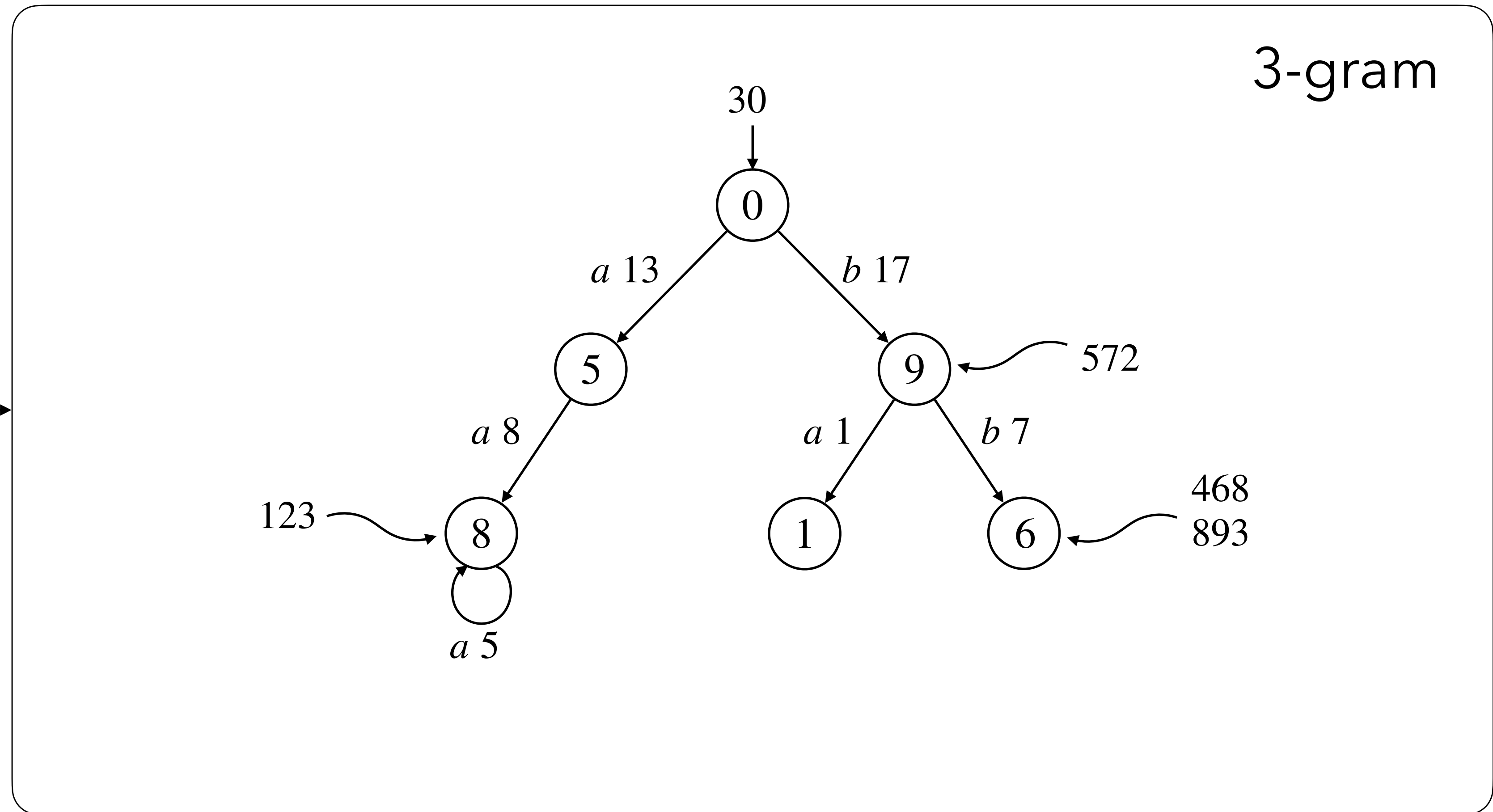
$\langle 123 : a \rangle$

$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	Lactic Acid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	Lactic Acid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

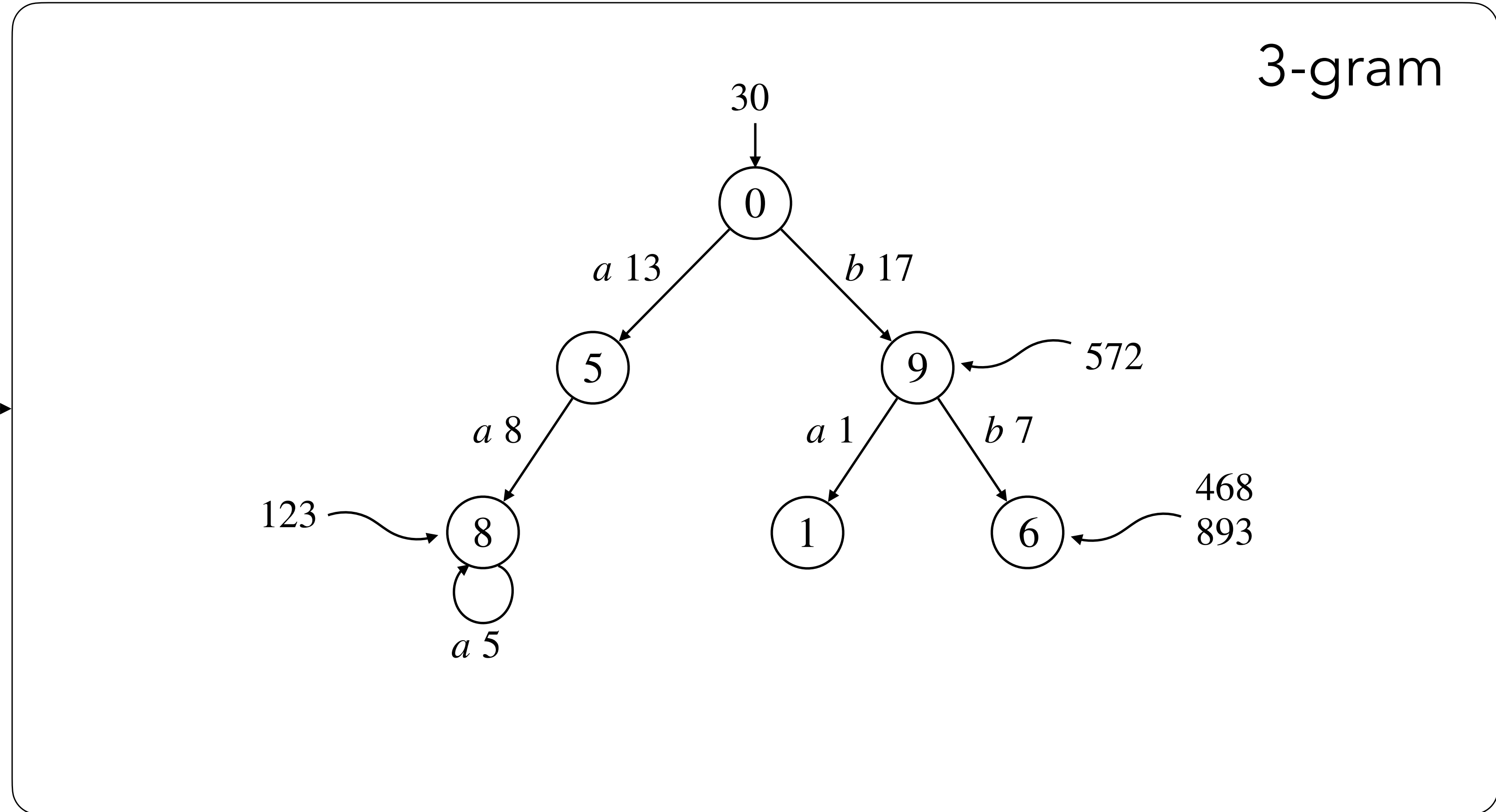
$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

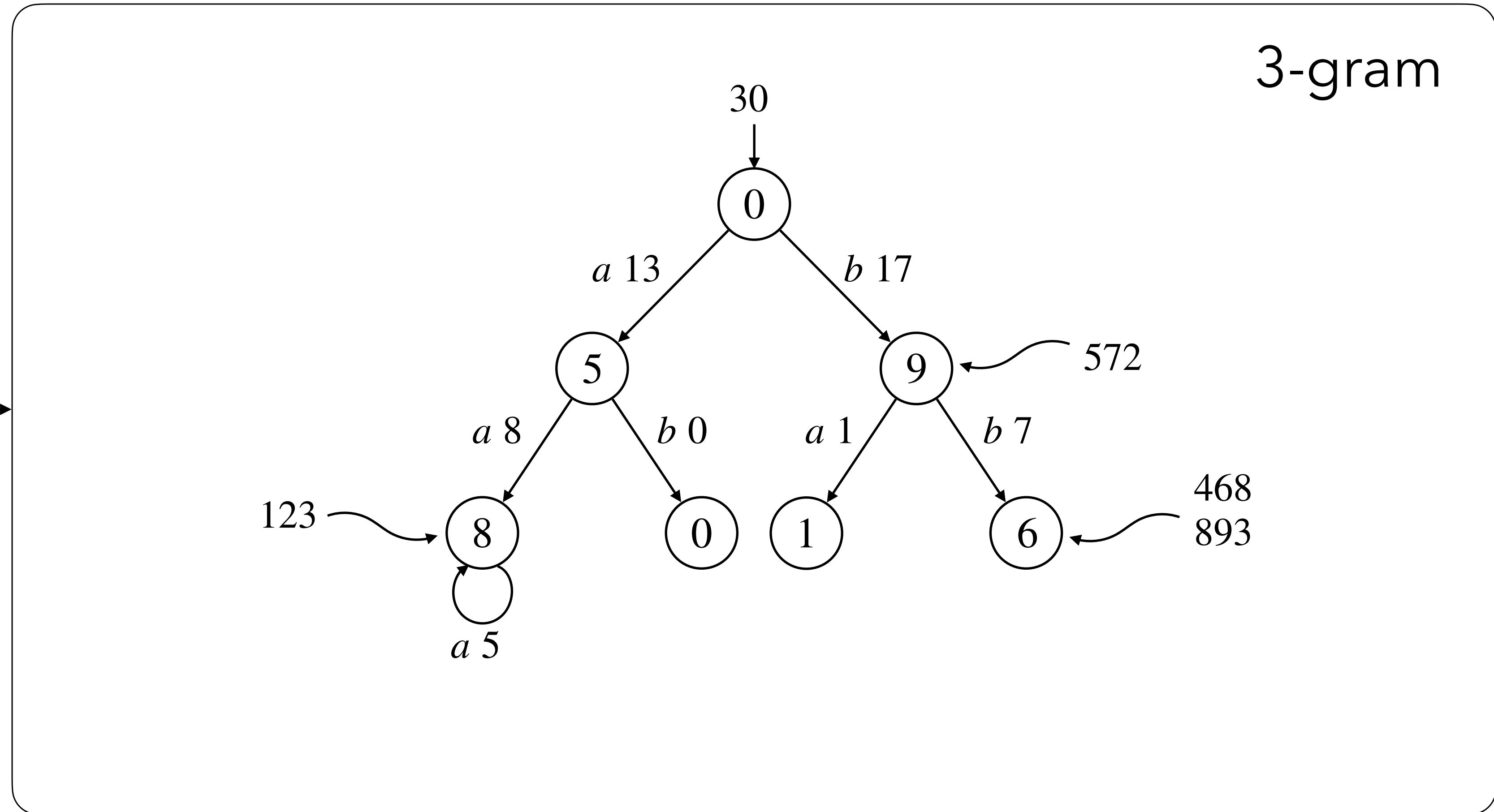
$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

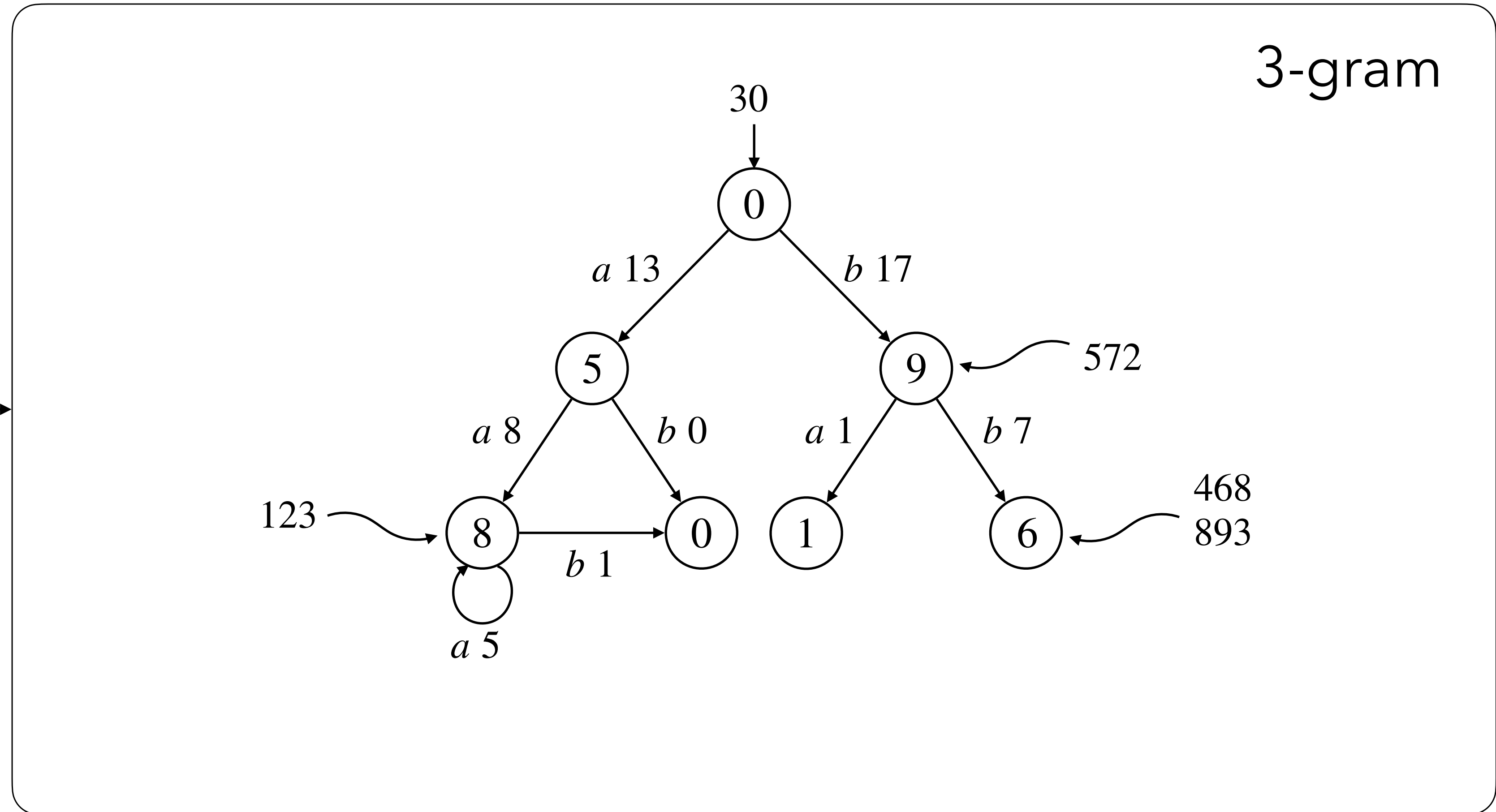
$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

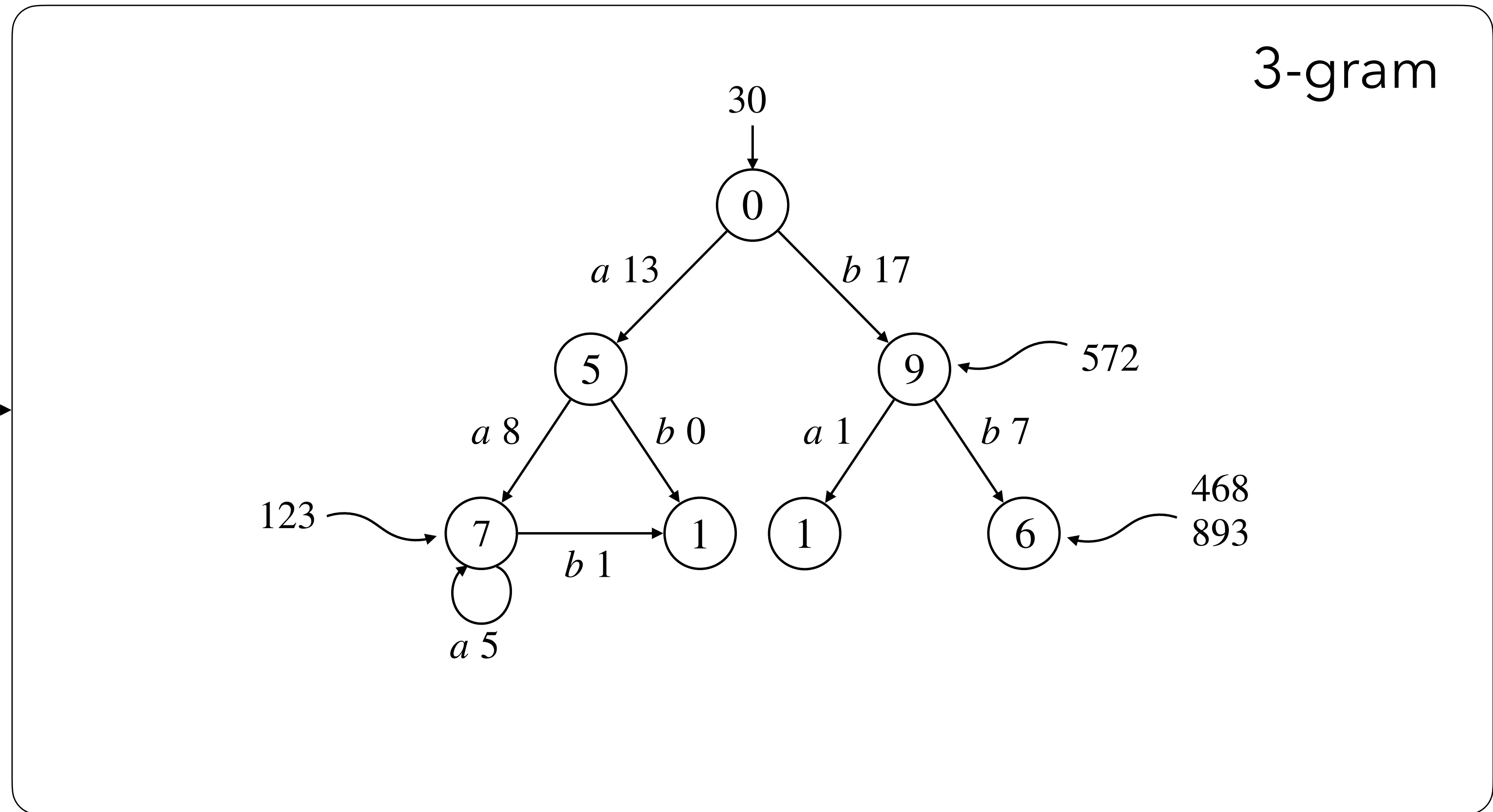
$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	Lactic Acid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	Lactic Acid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

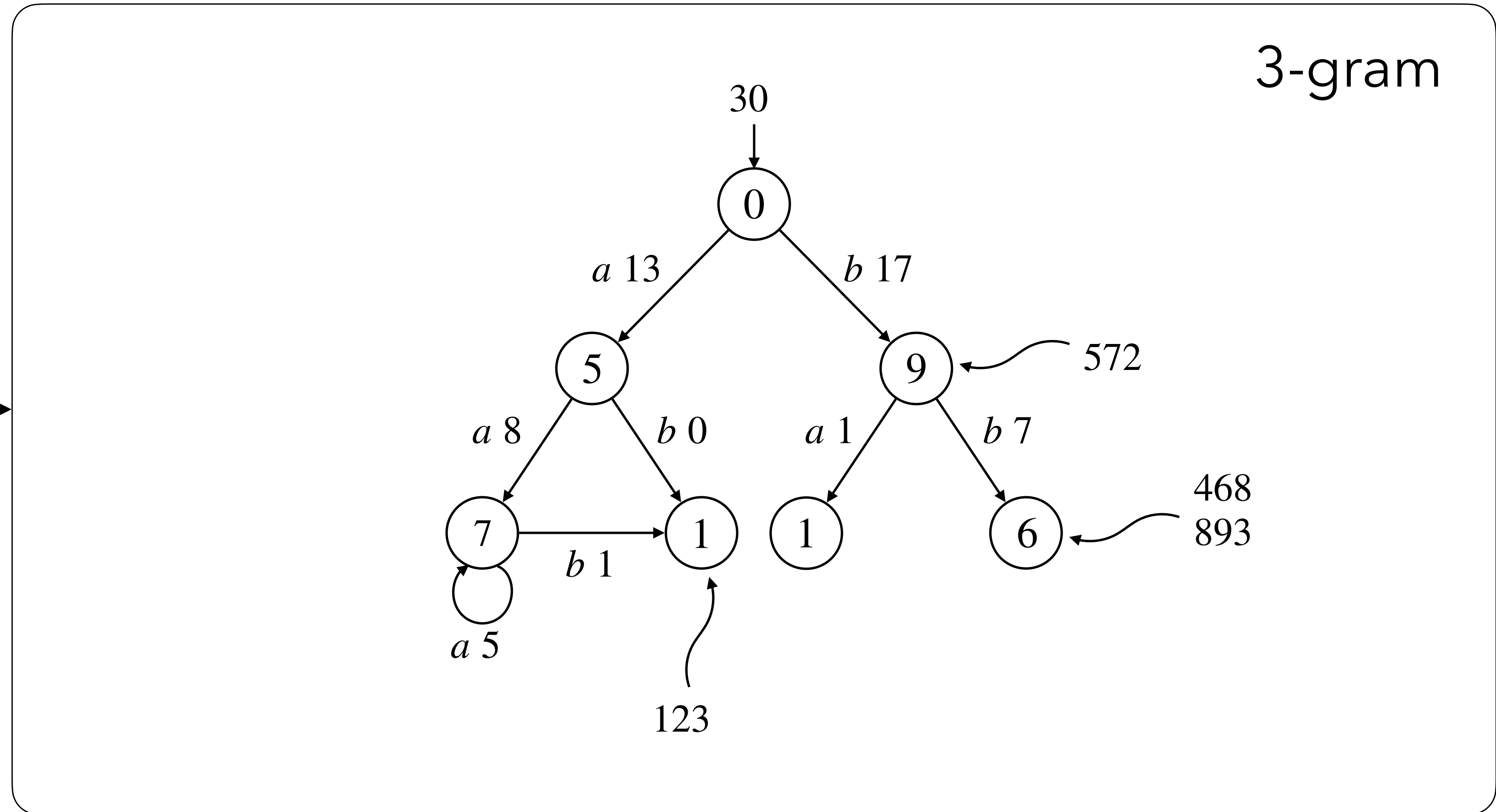
$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

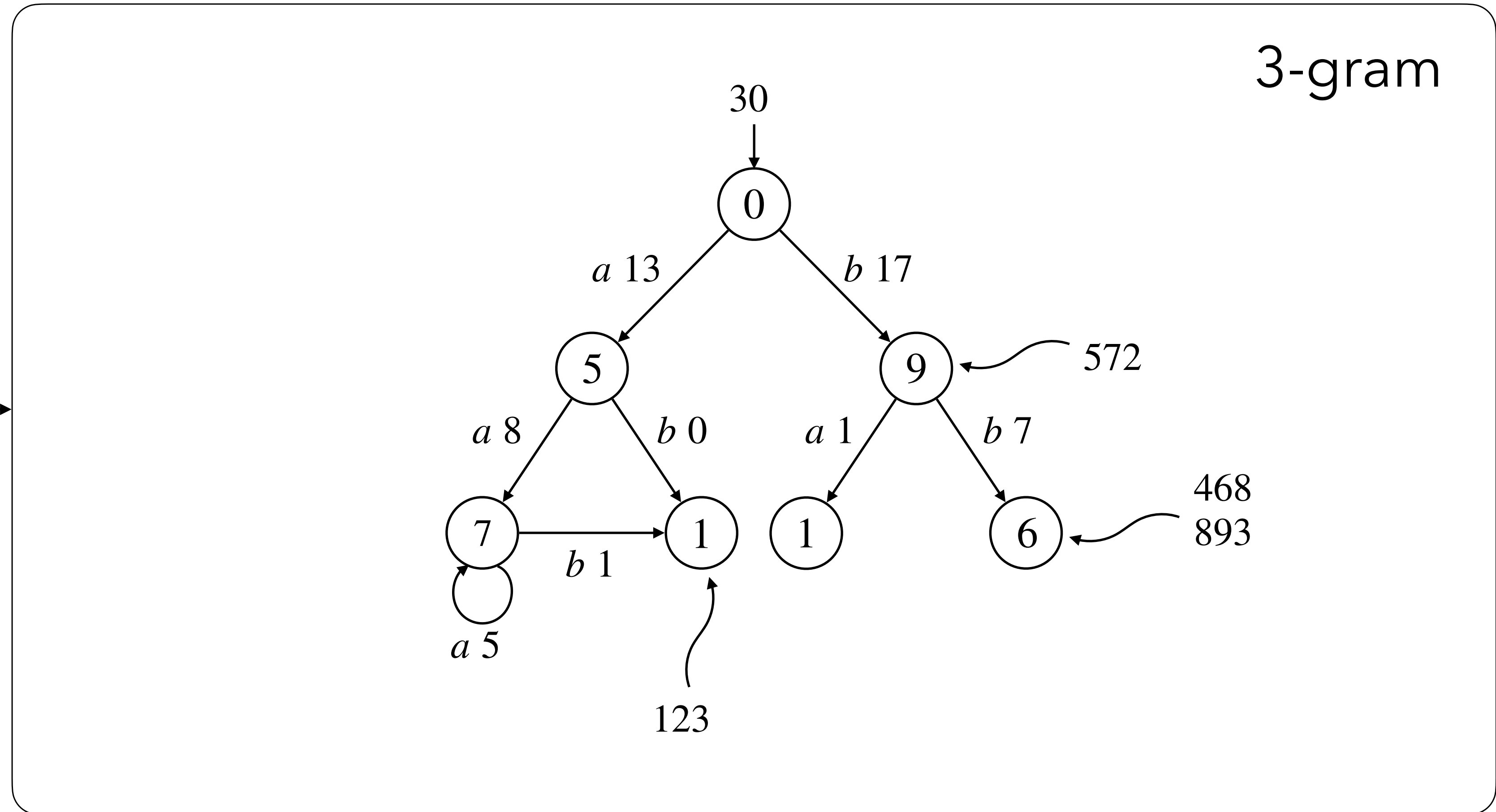
$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update
locally



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

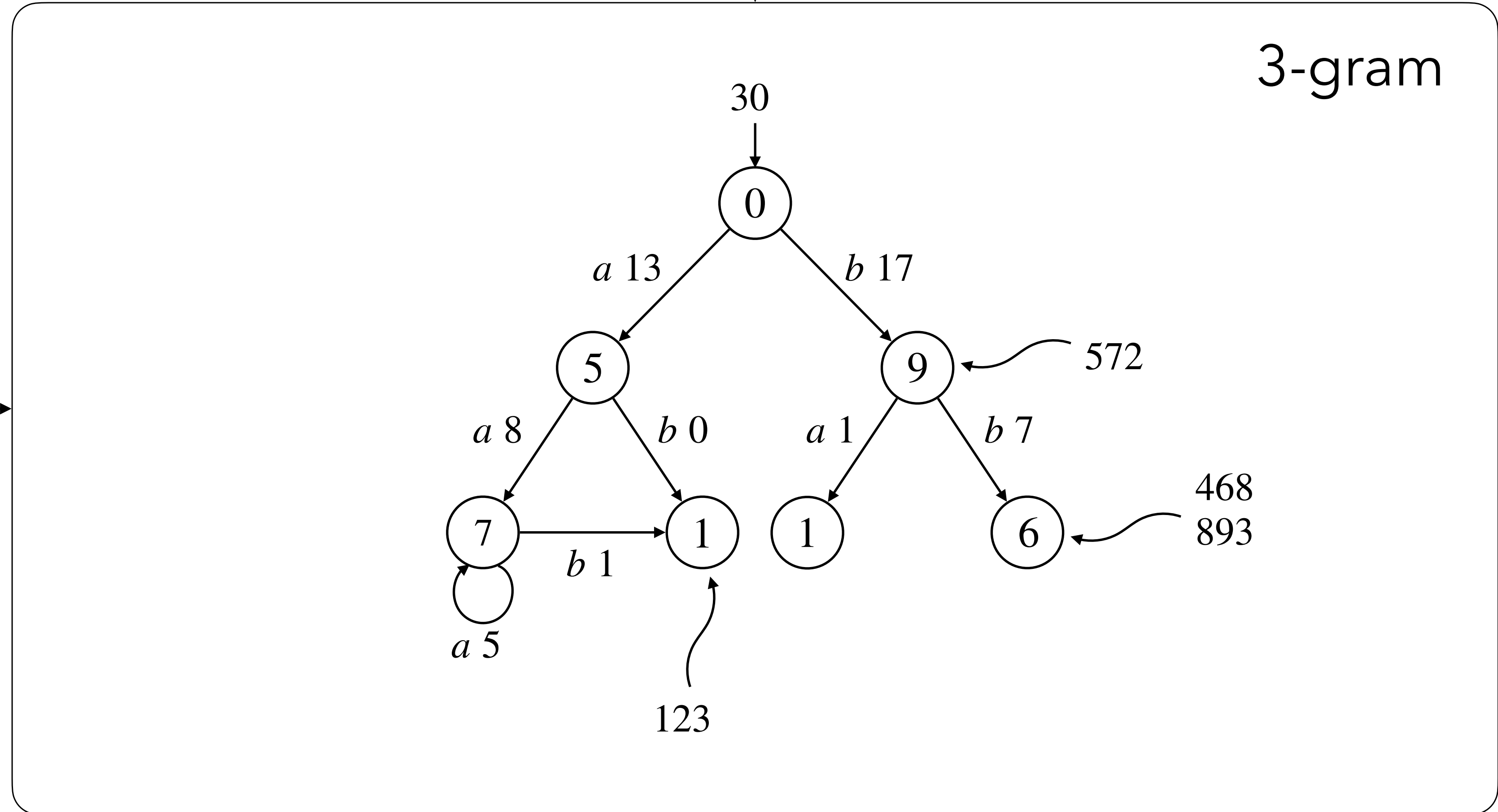
$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update
locally



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

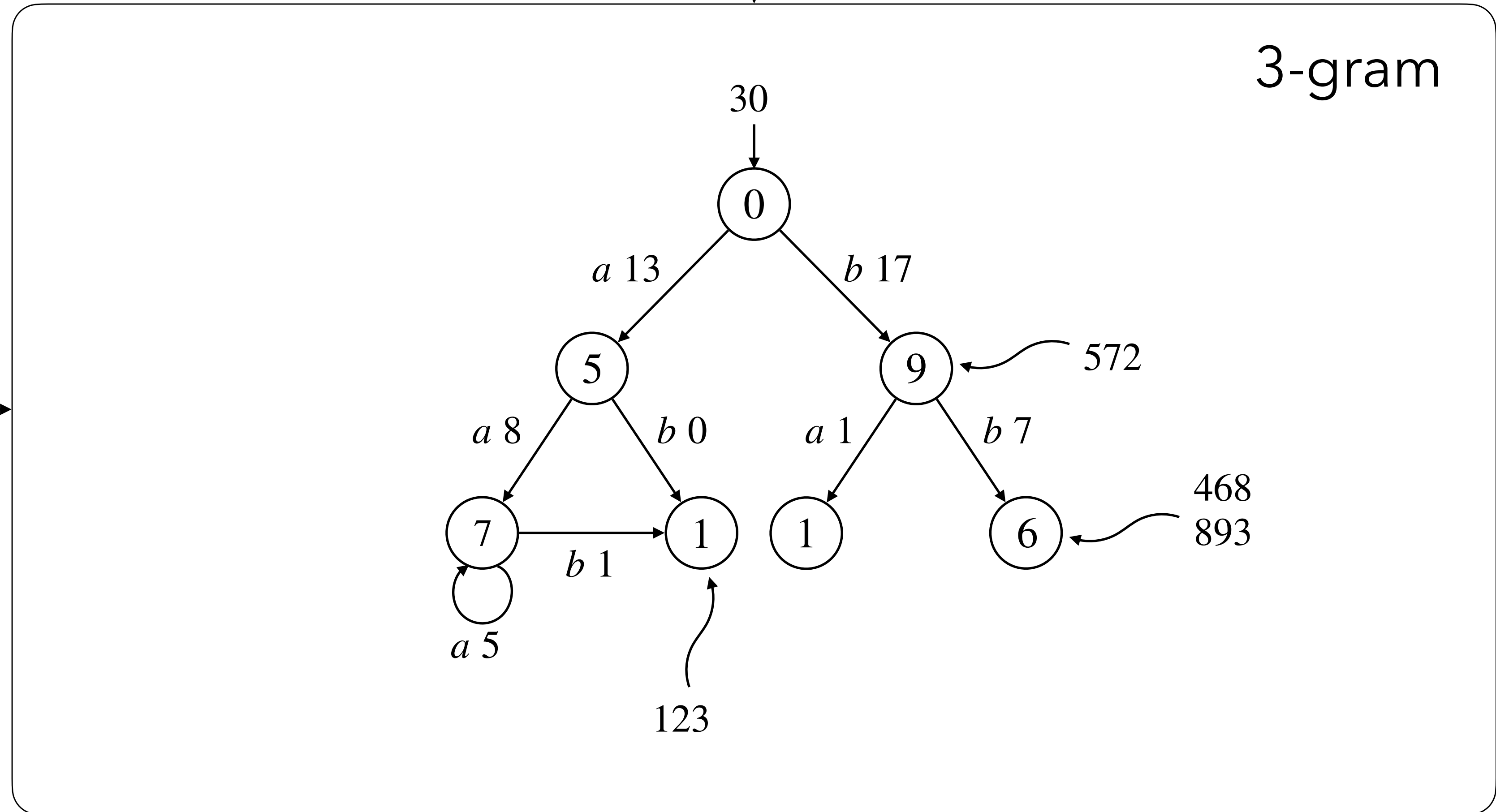
$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update
locally



Streaming learning: n-grams

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

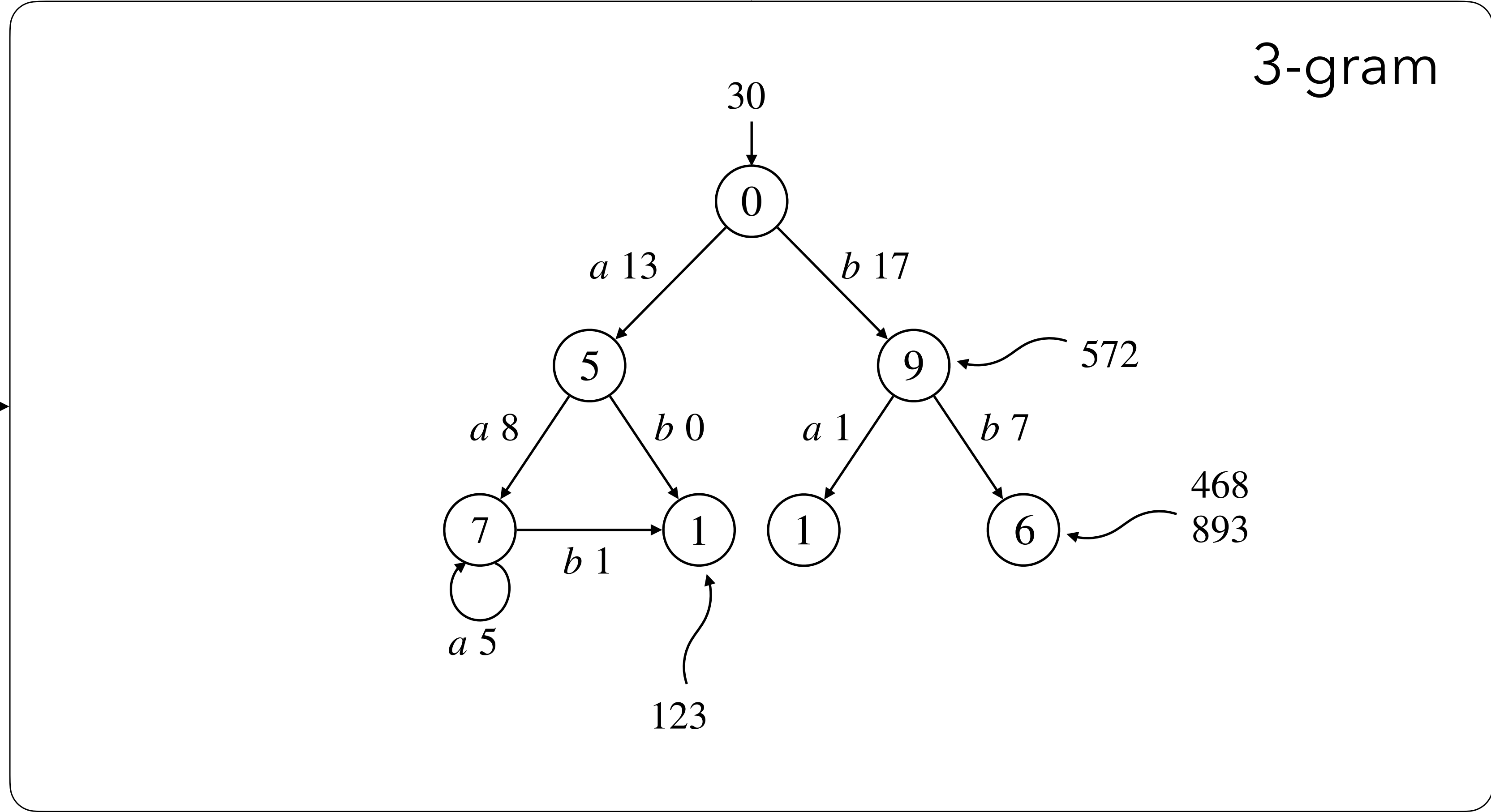
$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update
locally



- Back-off smoothing

stop

Streaming learning: Alergia

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	⋮
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

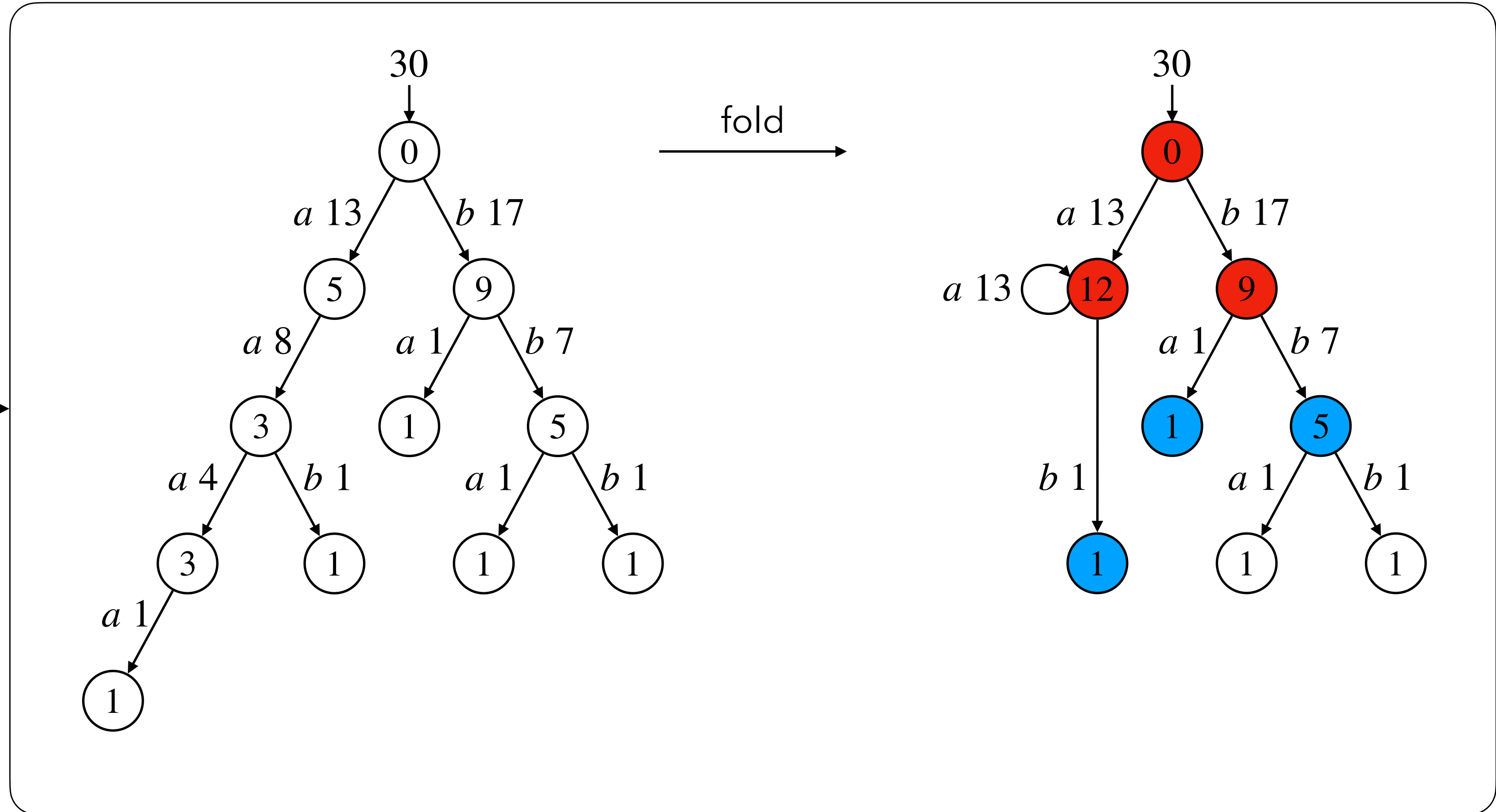
$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update



Streaming learning: Alergia

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

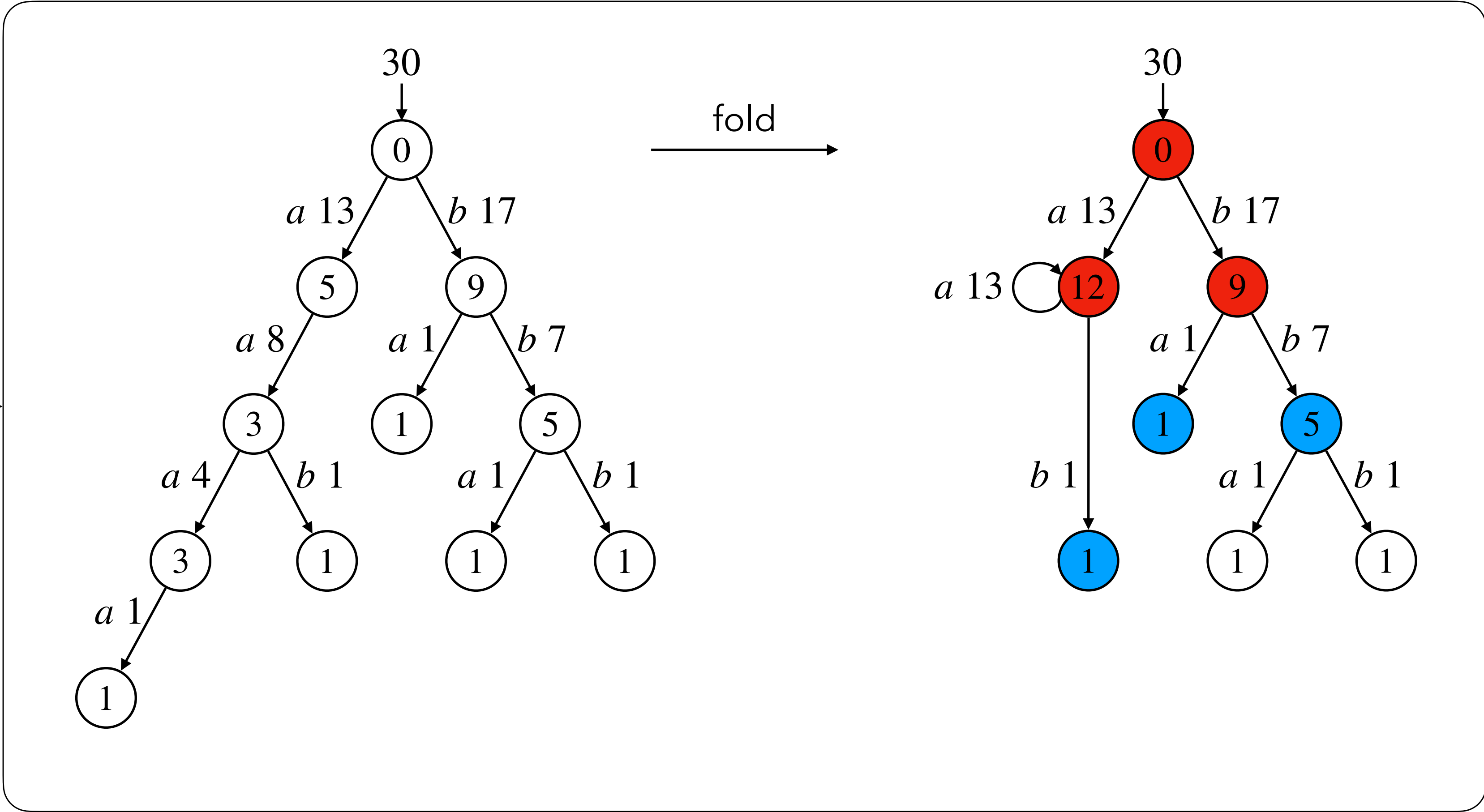
$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update
polynomial
update
time



Streaming learning: RNN/LSTM

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update



Streaming learning: RNN/LSTM

case:concept:name	concept:name
XJ	ER Registration
XJ	ER Triage
XJ	ER Sepsis Triage
XJ	LacticAcid
XJ	Leucocytes
XJ	CRP
XJ	IV Liquid
XJ	IV Antibiotics
XJ	Admission NC
XJ	Leucocytes
I	ER Registration
I	ER Triage
I	ER Sepsis Triage
I	IV Antibiotics
I	LacticAcid
I	CRP
I	Leucocytes
I	Admission NC
WEA	ER Registration
WEA	ER Triage
WEA	ER Sepsis Triage
WEA	CRP
WEA	Leucocytes
WEA	IV Liquid
WEA	IV Antibiotics
I	CRP
I	Leucocytes
OT	ER Registration
OT	ER Triage

$\langle 123 : a \rangle$

$\langle 572 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : a \rangle$

$\langle 893 : b \rangle$

$\langle 893 : b \rangle$

$\langle 468 : b \rangle$

$\langle 123 : b \rangle$

update

- Training iteration after each new event
- Small batch size n
- Include updated case sequence
- Add $n - 1$ sequences in round-robin fashion

More related work for streaming data

Balle & Castro & Gavaldà (2012). [Bootstrapping and Learning PDFAs in Data Streams.](#)

Burattin, Sperduti, van der Aalst (2014). [Control-flow discovery from event streams.](#)

Schmidt & Kramer (2014). [Online induction of probabilistic real-time automata.](#)

Baumgartner & Verwer (2023). [Learning state machines from data streams: A generic strategy and an improved heuristic.](#)

Ensemble Learning

model

model 1

model 2

model 3

model 4

Hard voting

model

model 1

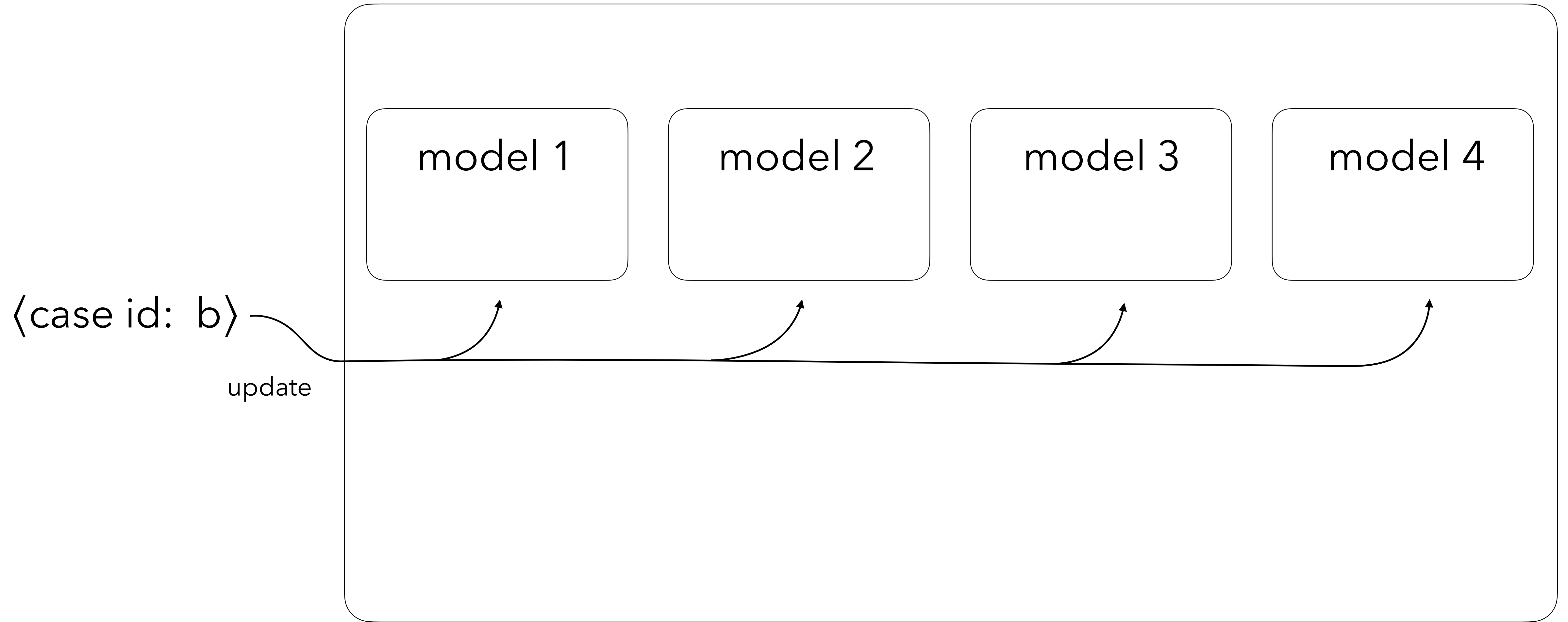
model 2

model 3

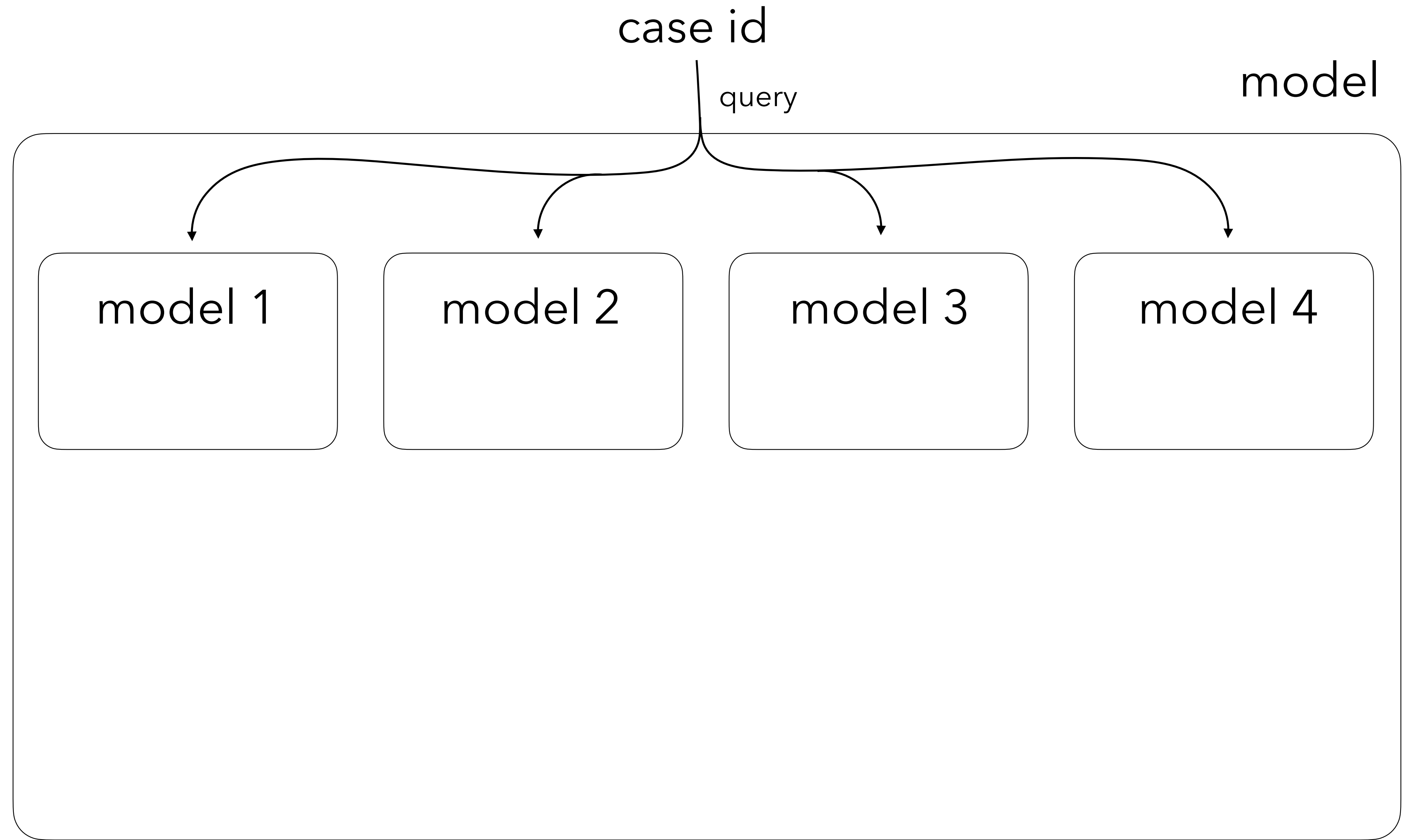
model 4

Hard voting

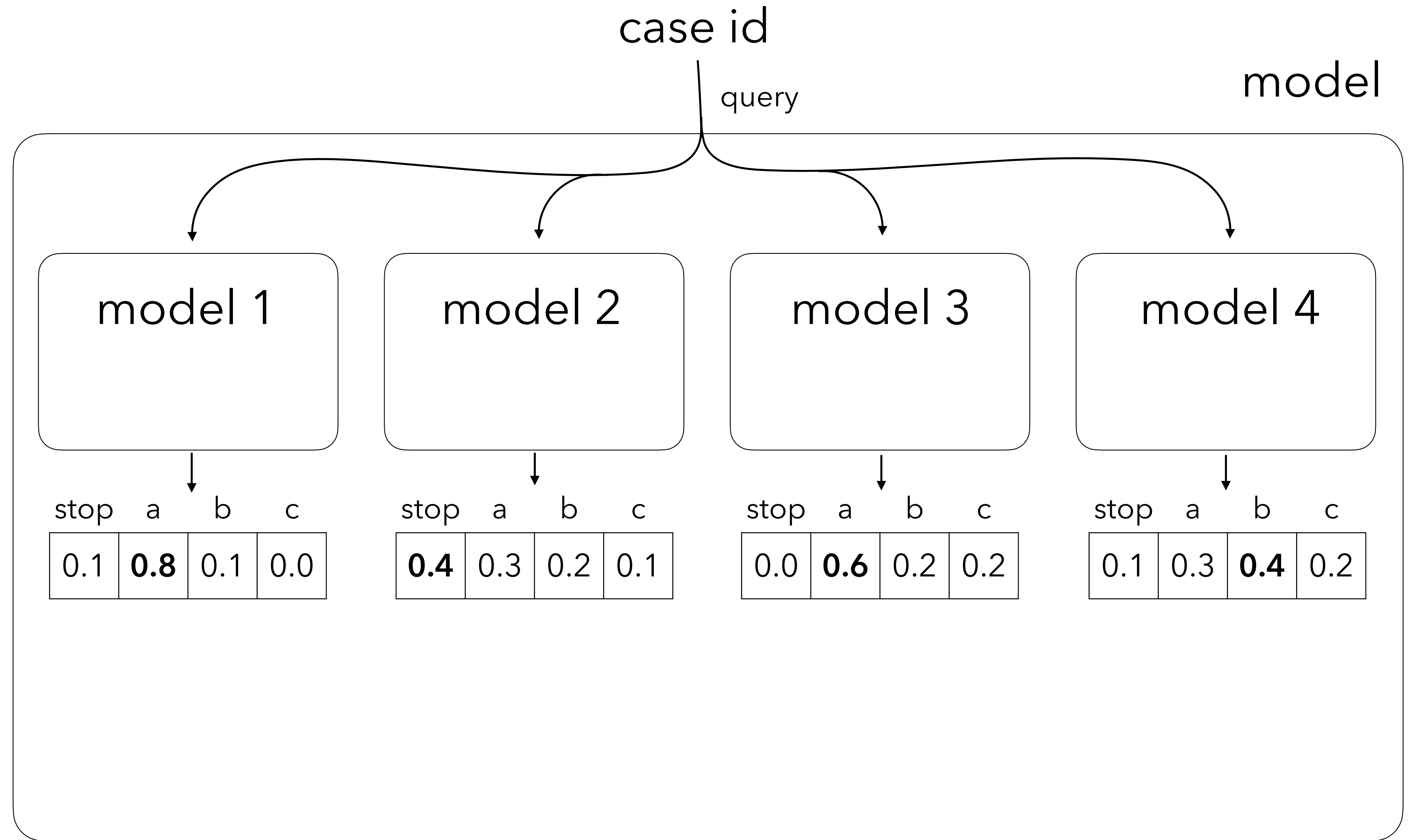
model



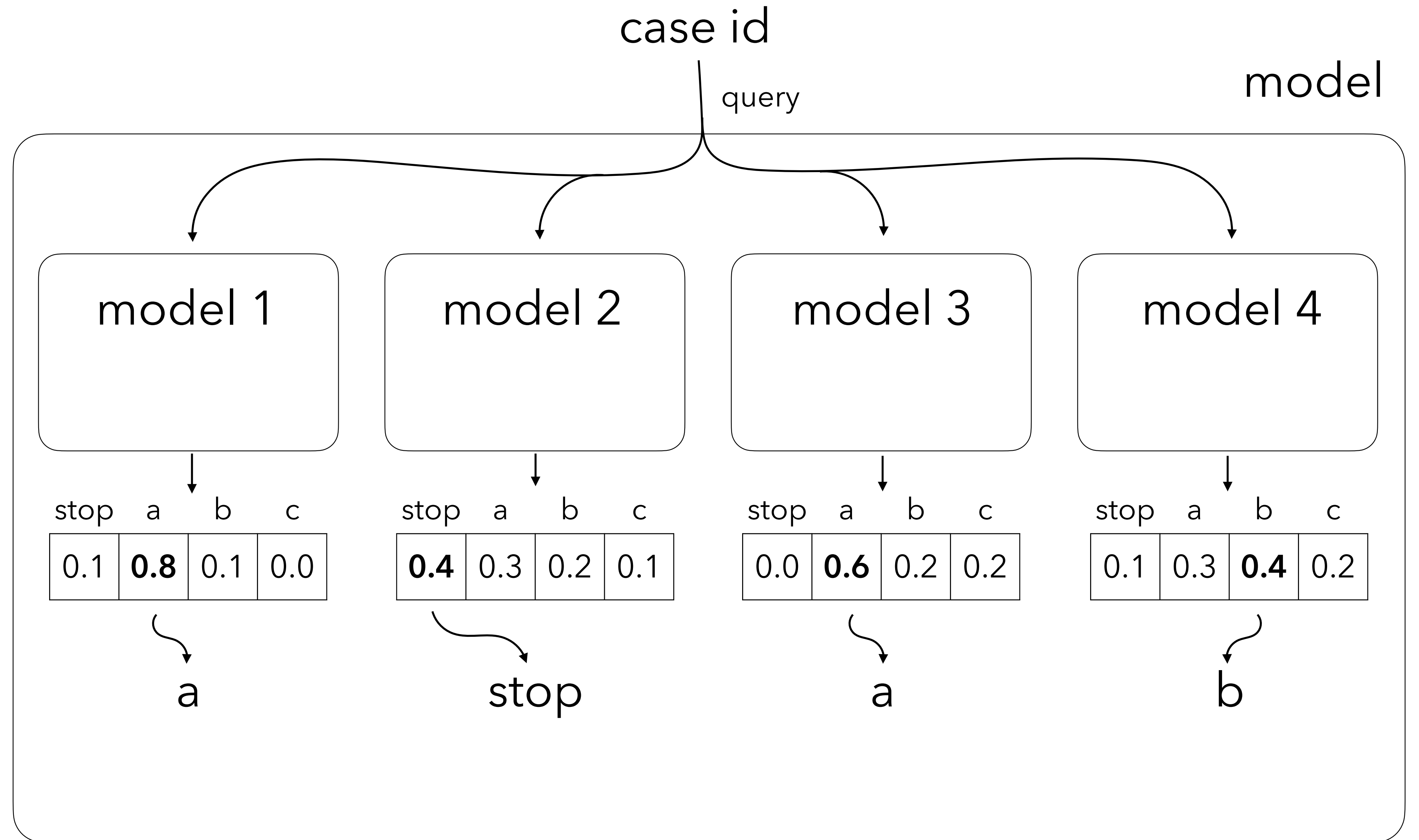
Hard voting



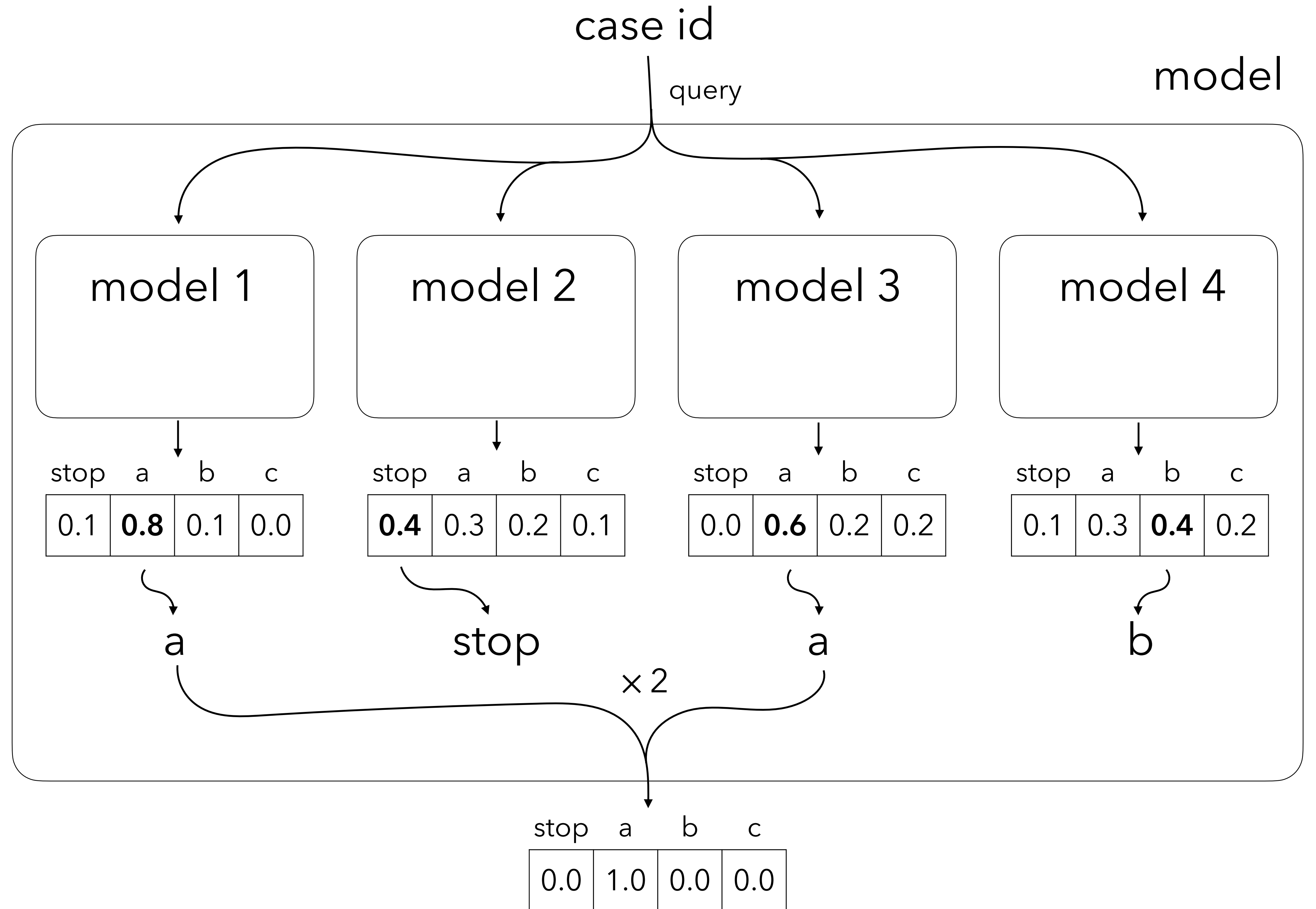
Hard voting



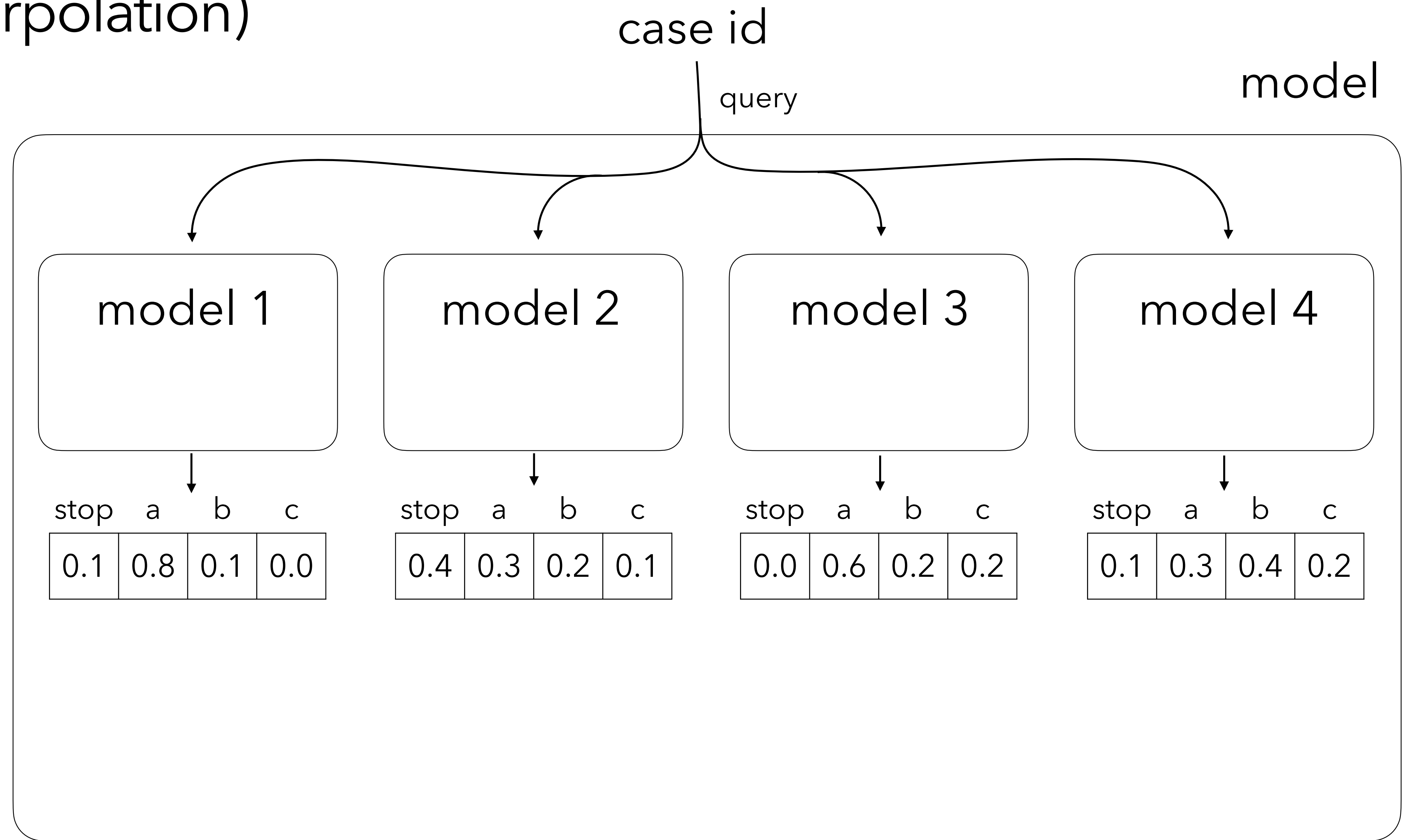
Hard voting



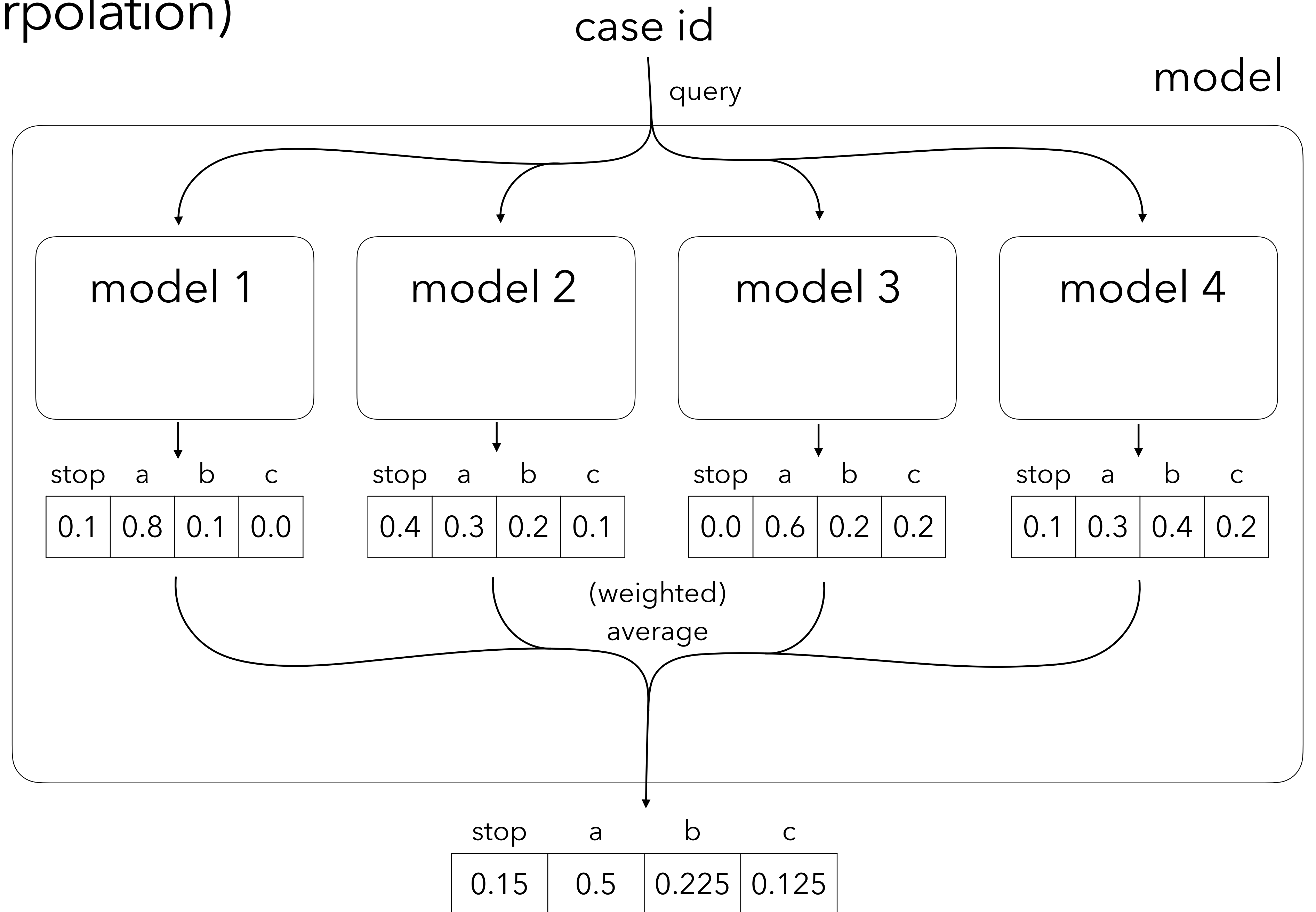
Hard voting



Soft voting (interpolation)



Soft voting (interpolation)



Adaptive voting

model

model 1

model 2

model 3

model 4

Adaptive voting

model

model 1

55%

model 2

60%

model 3

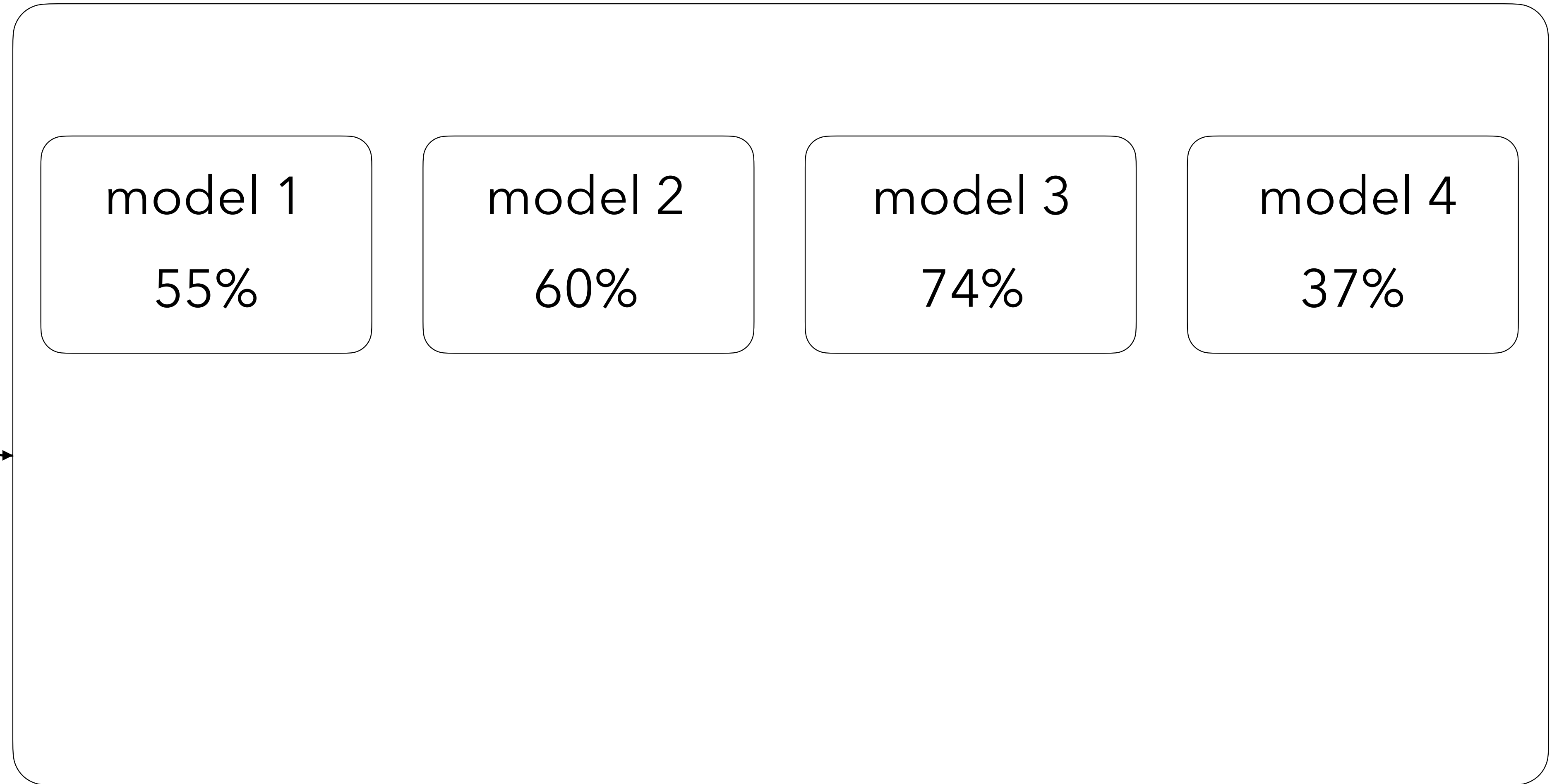
74%

model 4

37%

Adaptive voting

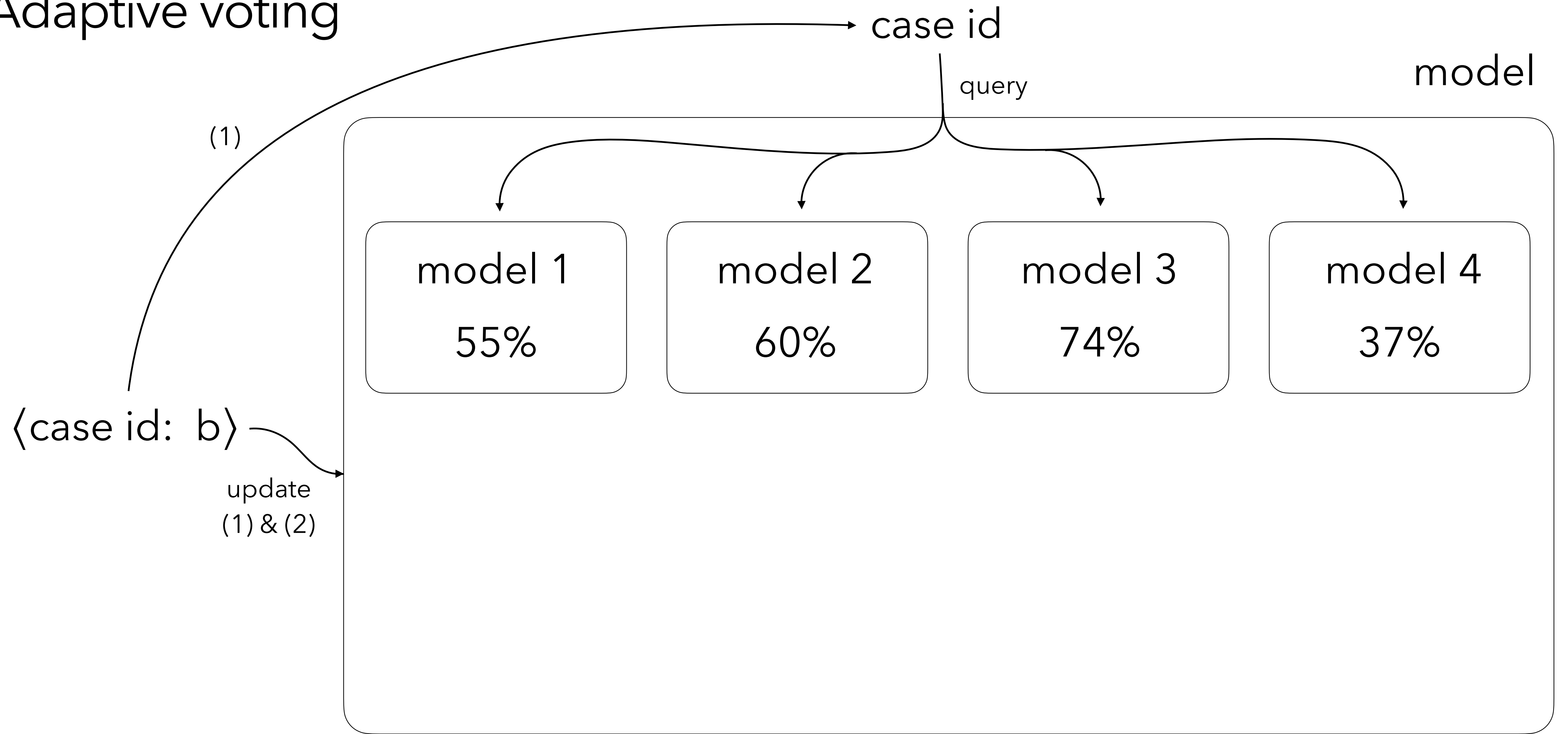
model



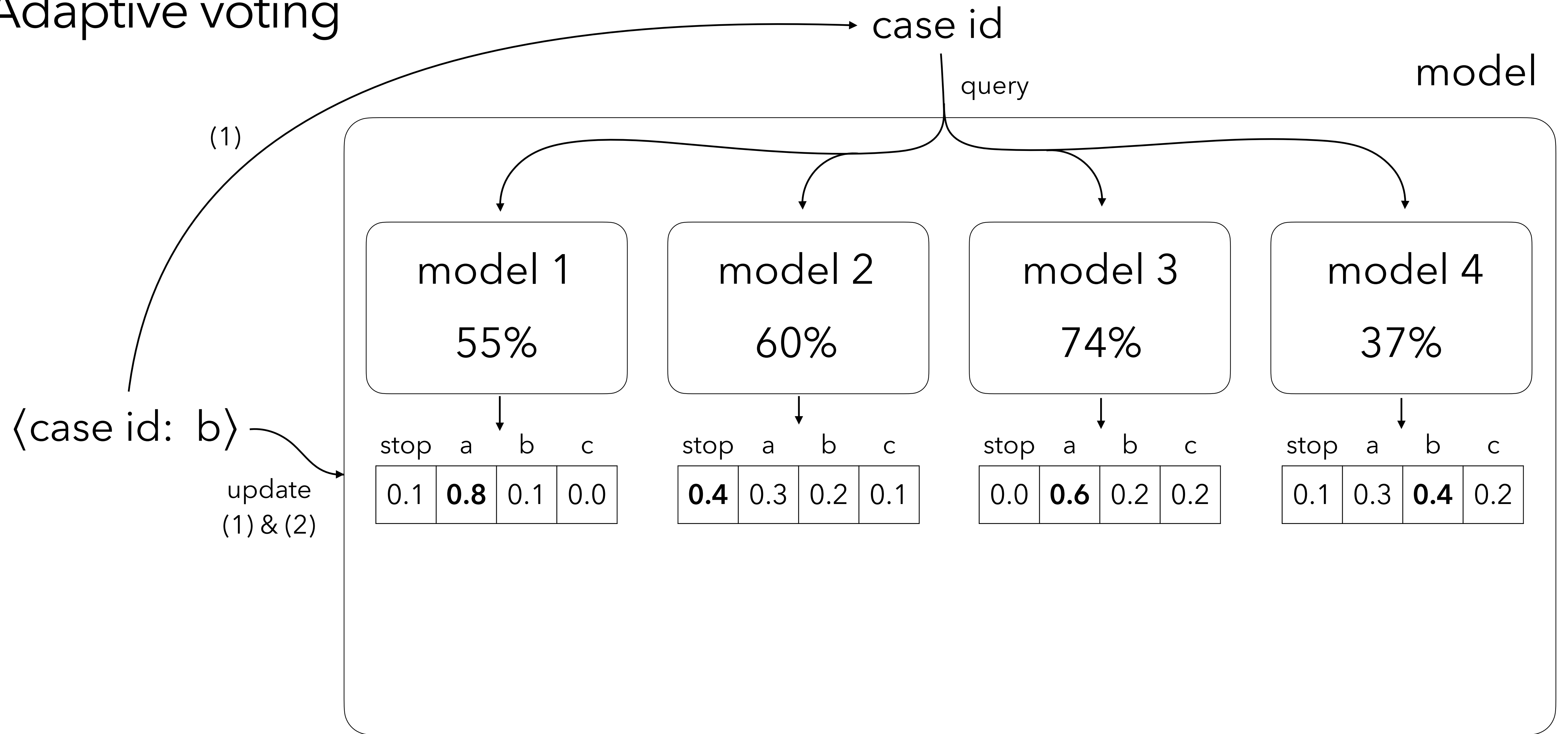
<case id: b>

update
(1) & (2)

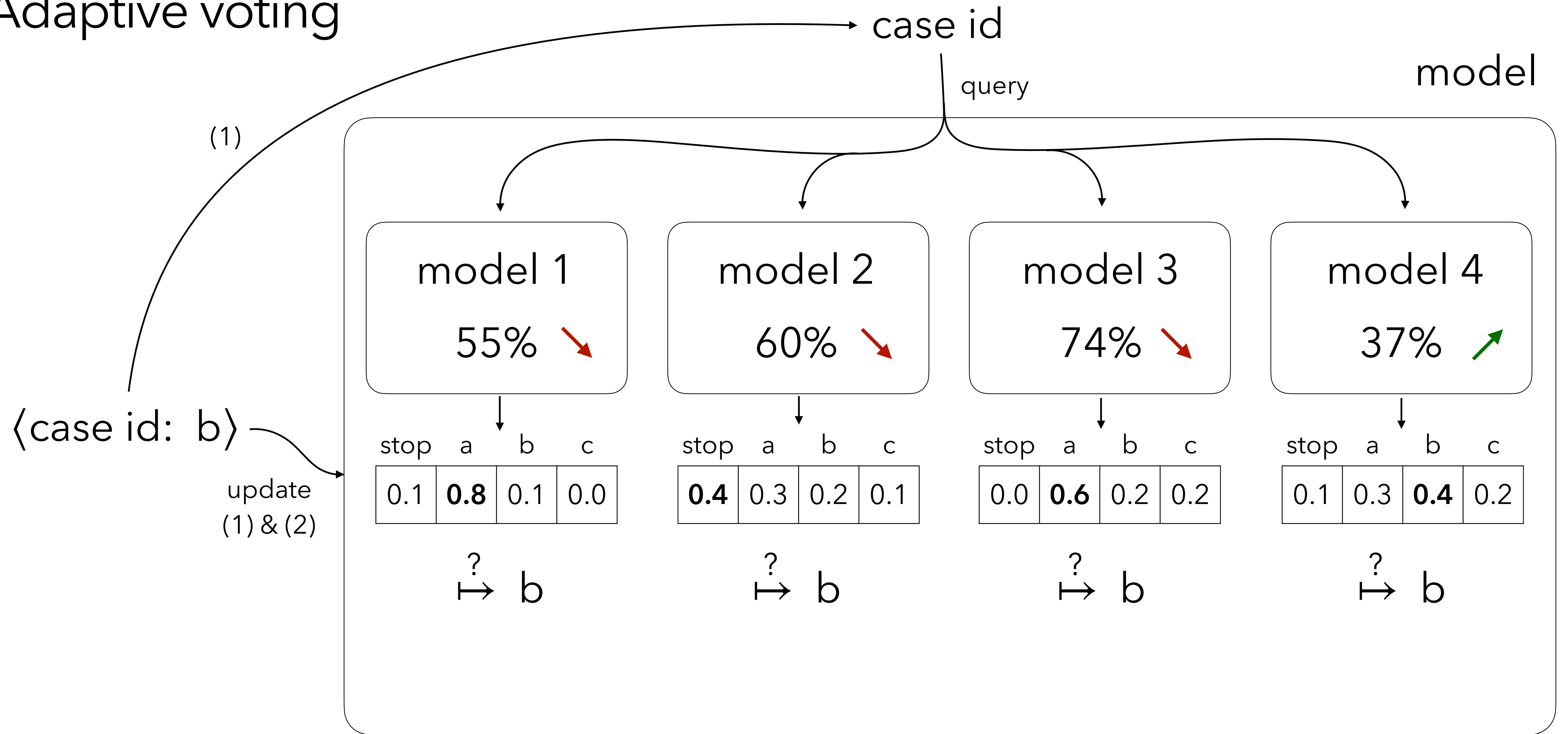
Adaptive voting



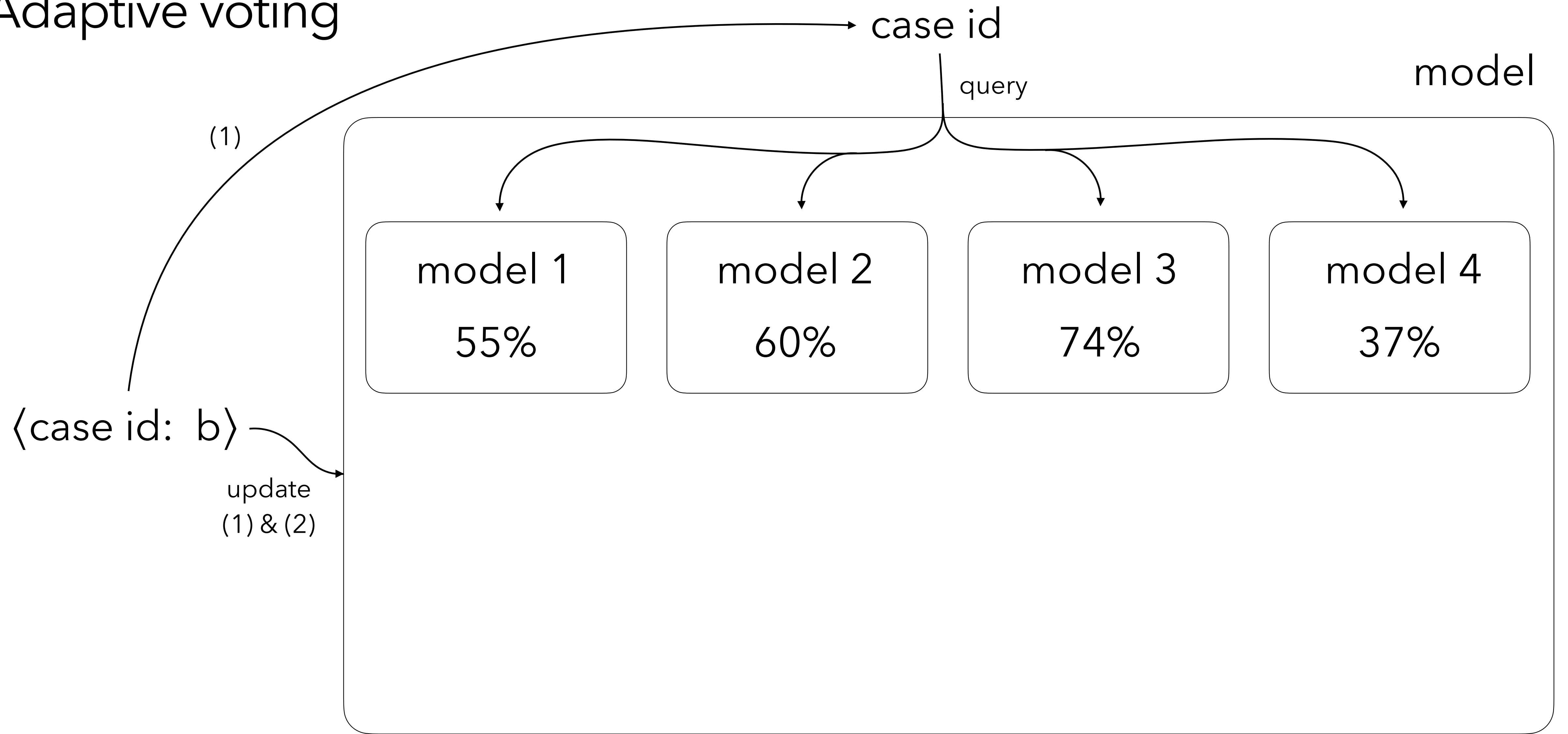
Adaptive voting



Adaptive voting

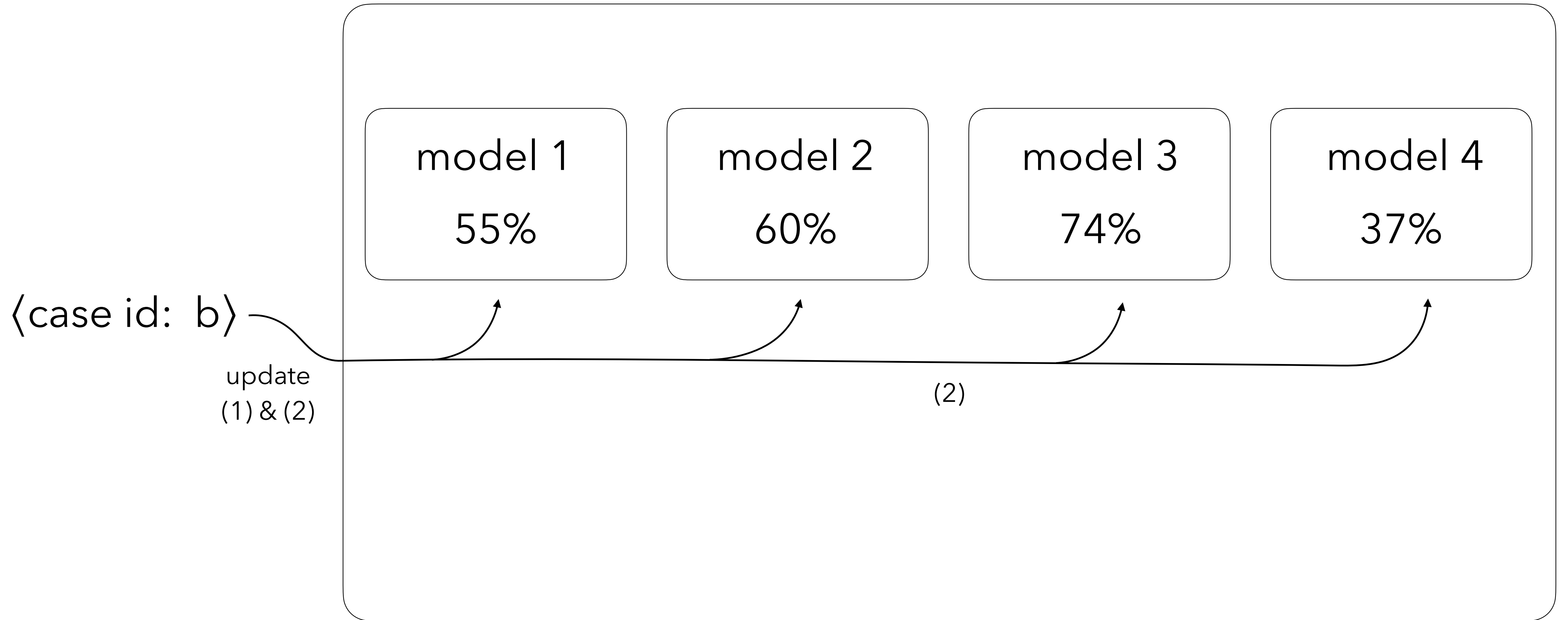


Adaptive voting



Adaptive voting

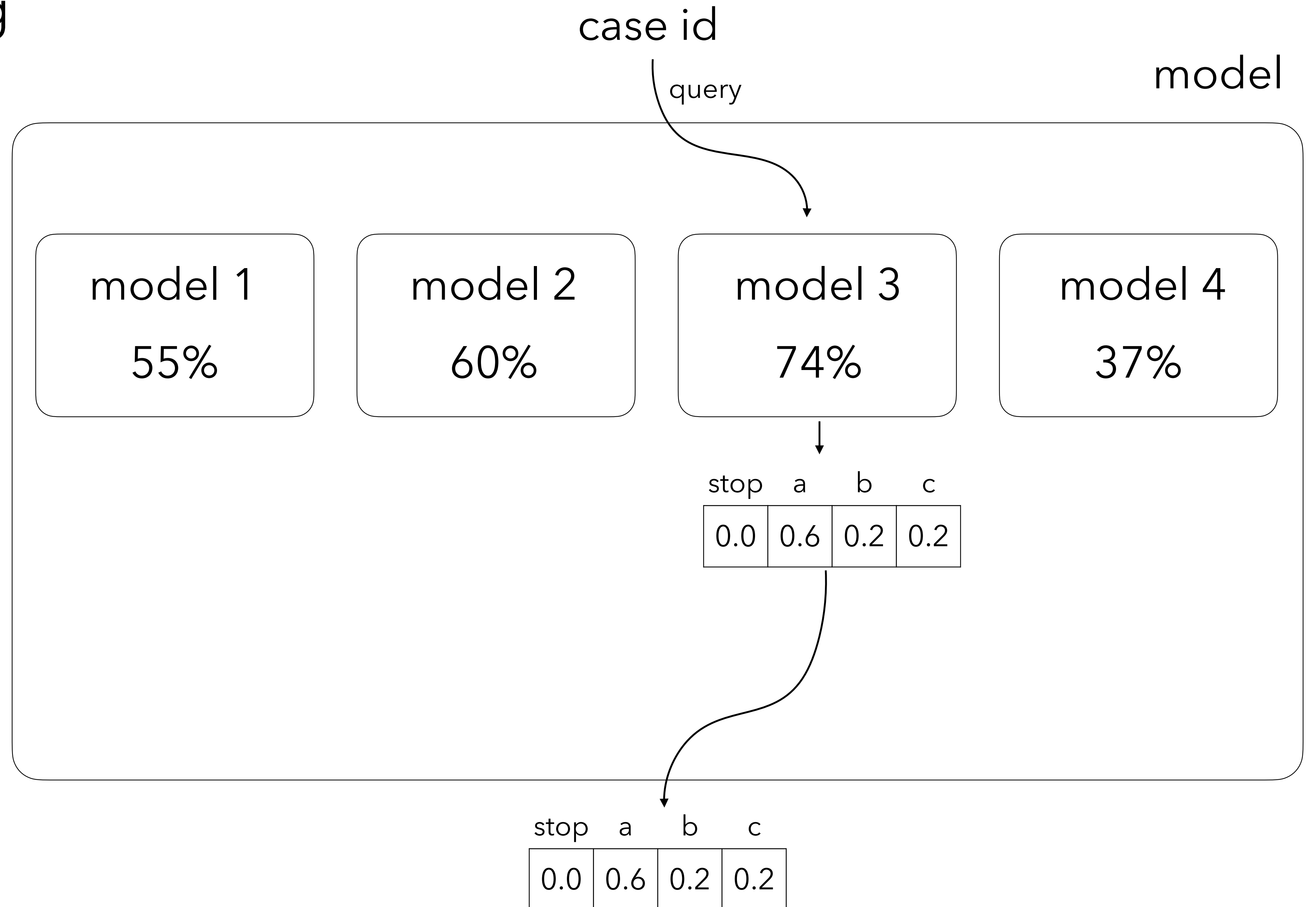
model



Adaptive voting



Adaptive voting



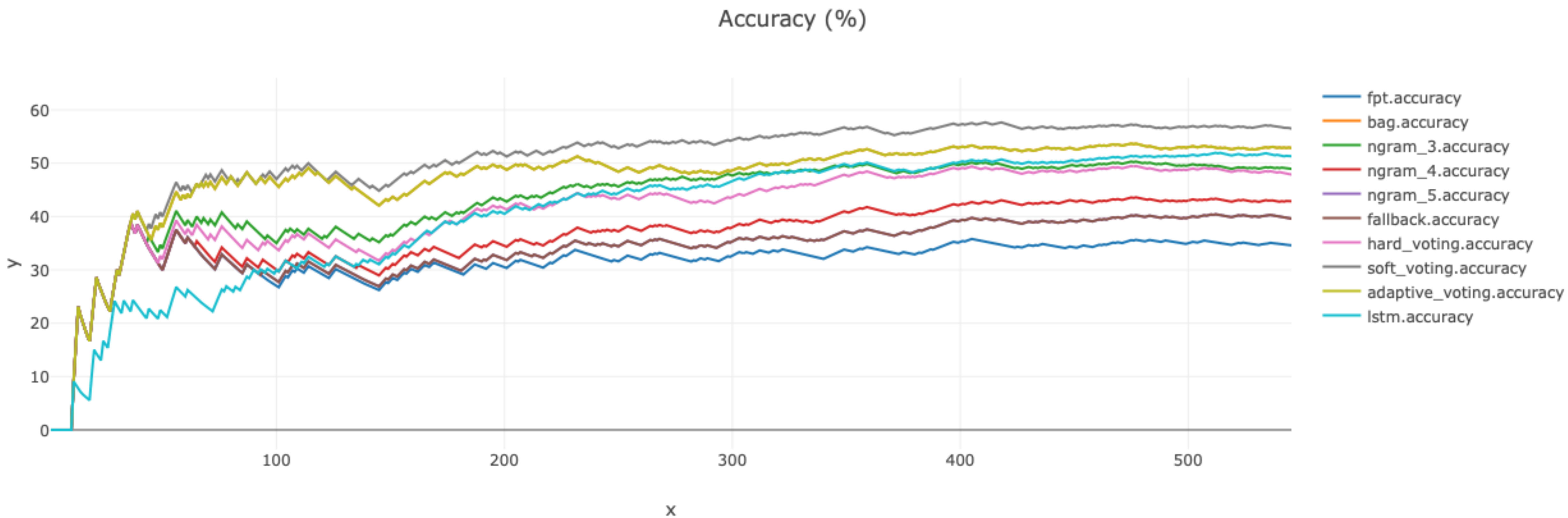
Experiments

Experiments

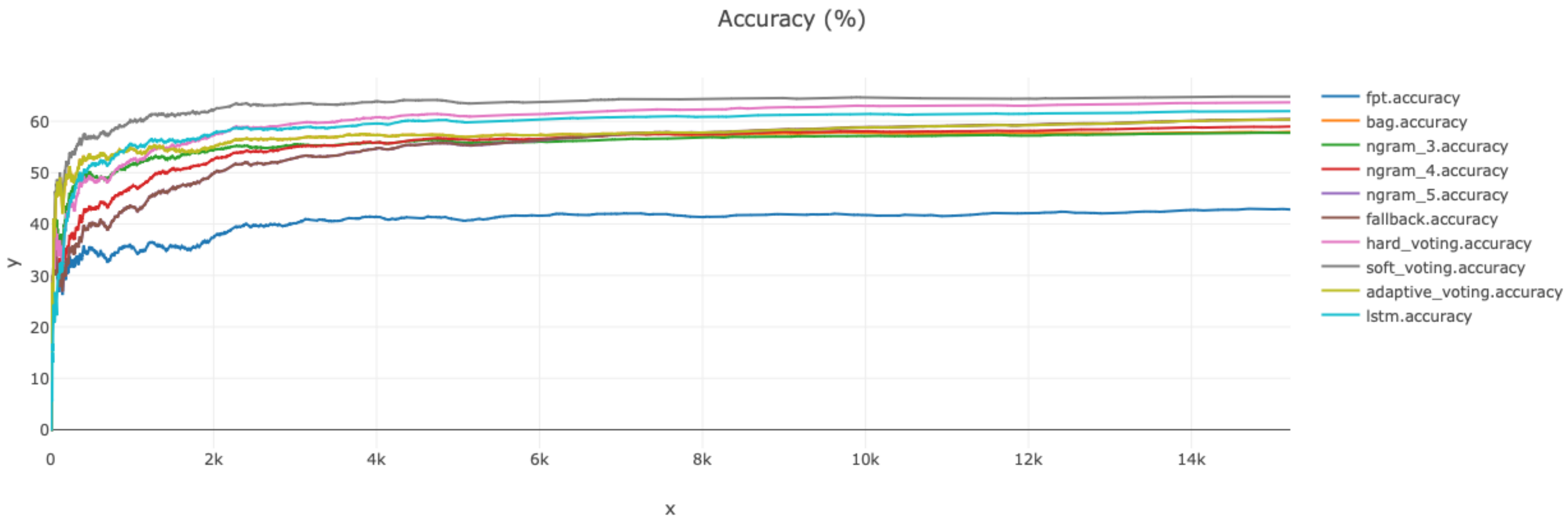
ensemble methods

	Sepsis Cases Events: 15214		BPI 2012 Events: 262200		BPI 2013 Events: 65533	
Batch	Accauracy (%)	# States	Accauracy (%)	# States	Accauracy (%)	# States
<input checked="" type="checkbox"/> fpt	43.26	4935	68.63	45277	54.01	13861
<input checked="" type="checkbox"/> bag	56.61	157	73.89	319	55.79	299
ngram_1	20.86	1	20.15	1	41.50	1
<input checked="" type="checkbox"/> ngram_3	57.78	127	81.70	147	70.42	92
<input checked="" type="checkbox"/> ngram_4	59.76	472	83.66	420	70.71	410
<input checked="" type="checkbox"/> ngram_5	61.81	1250	84.82	876	70.39	1267
ngram_6	59.85	2625	85.57	1552	69.79	2987
alergia	50.95	47	75.25	136	55.72	32
LSTM	64.00		85.71		70.86	
hard voting	64.09		85.43		70.81	
soft voting	64.20		85.55		70.99	
Streaming	Accauracy (%)	Mean Latency (ms)	Accauracy (%)	Mean Latency (ms)	Accauracy (%)	Mean Latency (ms)
fpt	42.84	0.04	66.39	0.06	51.09	0.04
bag	57.95	0.09	75.86	0.06	52.68	0.04
ngram_1	22.21	0.13	20.91	0.06	46.14	0.06
ngram_3	57.76	0.08	81.37	0.06	67.17	0.04
ngram_4	58.98	0.08	83.34	0.07	67.19	0.04
ngram_5	60.37	0.13	84.48	0.06	66.49	0.04
ngram_6	57.80	0.12	85.15	0.06	65.22	0.05
LSTM	61.66	6.41	84.34	4.46	66.33	7.00
hard voting	63.68	0.28	84.97	0.58	67.23	0.17
soft voting	64.80	0.17	85.20	0.11	67.46	0.07
adaptive voting	60.37	0.28	85.15	0.23	67.19	0.15

Experiments



Experiments



A Python streaming library

```
soft_voting = StreamingActivityPredictor(  
    strategy=SoftVoting(  
        models=[  
            BasicMiner(algorithm=Bag()),  
            BasicMiner(algorithm=FrequencyPrefixTree()),  
            BasicMiner(algorithm=NGram(window_length=2)),  
            BasicMiner(algorithm=NGram(window_length=3)),  
            BasicMiner(algorithm=NGram(window_length=4)),  
        ],  
        config=config,  
    )  
)
```

```
sponge = (  
    streamer  
    * ls.KeyFilter(keys=["case_id", "action"])  
    * AddStartSymbol()  
    * (  
        (fpt * Evaluation("fpt"))  
        | (bag * Evaluation("bag"))  
        | (ngram_1 * Evaluation("ngram_1"))  
        | (ngram_3 * Evaluation("ngram_3"))  
        | (ngram_5 * Evaluation("ngram_5"))  
        | (fallback * Evaluation("fallback"))  
        | (hard_voting * Evaluation("hard_voting"))  
        | (soft_voting * Evaluation("soft_voting"))  
        | (adaptive_voting * Evaluation("adaptive_voting"))  
        | (lstm * Evaluation("lstm"))  
    )  
    * ls.ToSingleStream(flatten=True)  
    * ls.AddIndex(key="index")  
    * ls.KeyFilter(keys=accuracy_list)  
    * (dashboard.Plot("Accuracy (%)", x="index", y=accuracy_list))  
)  
  
sponge.start()
```

A Python streaming library

```
soft_voting = StreamingActivityPredictor(  
    strategy=SoftVoting(  
        models=[  
            BasicMiner(algorithm=Bag()),  
            BasicMiner(algorithm=FrequencyPrefixTree()),  
            BasicMiner(algorithm=NGram(window_length=2)),  
            BasicMiner(algorithm=NGram(window_length=3)),  
            BasicMiner(algorithm=NGram(window_length=4)),  
        ],  
        config=config,  
    )  
)
```

```
sponge = (  
    streamer  
    * ls.KeyFilter(keys=["case_id", "action"])  
    * AddStartSymbol()  
    * (  
        (fpt * Evaluation("fpt"))  
        | (bag * Evaluation("bag"))  
        | (ngram_1 * Evaluation("ngram_1"))  
        | (ngram_3 * Evaluation("ngram_3"))  
        | (ngram_5 * Evaluation("ngram_5"))  
        | (fallback * Evaluation("fallback"))  
        | (hard_voting * Evaluation("hard_voting"))  
        | (soft_voting * Evaluation("soft_voting"))  
        | (adaptive_voting * Evaluation("adaptive_voting"))  
        | (lstm * Evaluation("lstm"))  
    )  
    * ls.ToSingleStream(flatten=True)  
    * ls.AddIndex(key="index")  
    * ls.KeyFilter(keys=accuracy_list)  
    * (dashboard.Plot("Accuracy (%)", x="index", y=accuracy_list))  
)  
  
sponge.start()
```

Logicsponge 🧽

<https://github.com/logicsponge>

Conclusion

Conclusion

- Process data is inherently "n-gram"
- Ensemble method boost small models

Conclusion

- Process data is inherently "n-gram"
- Ensemble method boost small models
- Falback algorithms and bisimilarity distance
- Beyond case-level predictions: Object-centric event data

Ghahfarokhi, Park, Berti, van der Aalst (2021): [OCEL: A Standard for Object-Centric Event Logs](#).

Bojanczyk, David, Muscholl, Schwentick, Segoufin (2011): [Two-variable logic on data words](#).

Decker, Habermehl, Leucker, Thoma (2014): [Learning Transparent Data Automata](#).