LA-UR-21-21679

Approved for public release; distribution is unlimited.

Title: Proliferation Monitoring with Hidden Markov Models

Author(s): Hollis, Andrew
Tompkins, George
Wilson, Alyson
Smith, Ralph

Intended for: Next-Gen AI for Proliferation Detection: Domain Aware Methods (Virtual Workshop)

Issued: 2021-02-22

Proliferation Monitoring with Hidden Markov Models

Andrew Hollis, George Tompkins, Alyson Wilson, and Ralph Smith

January, 2021

Proliferation Monitoring

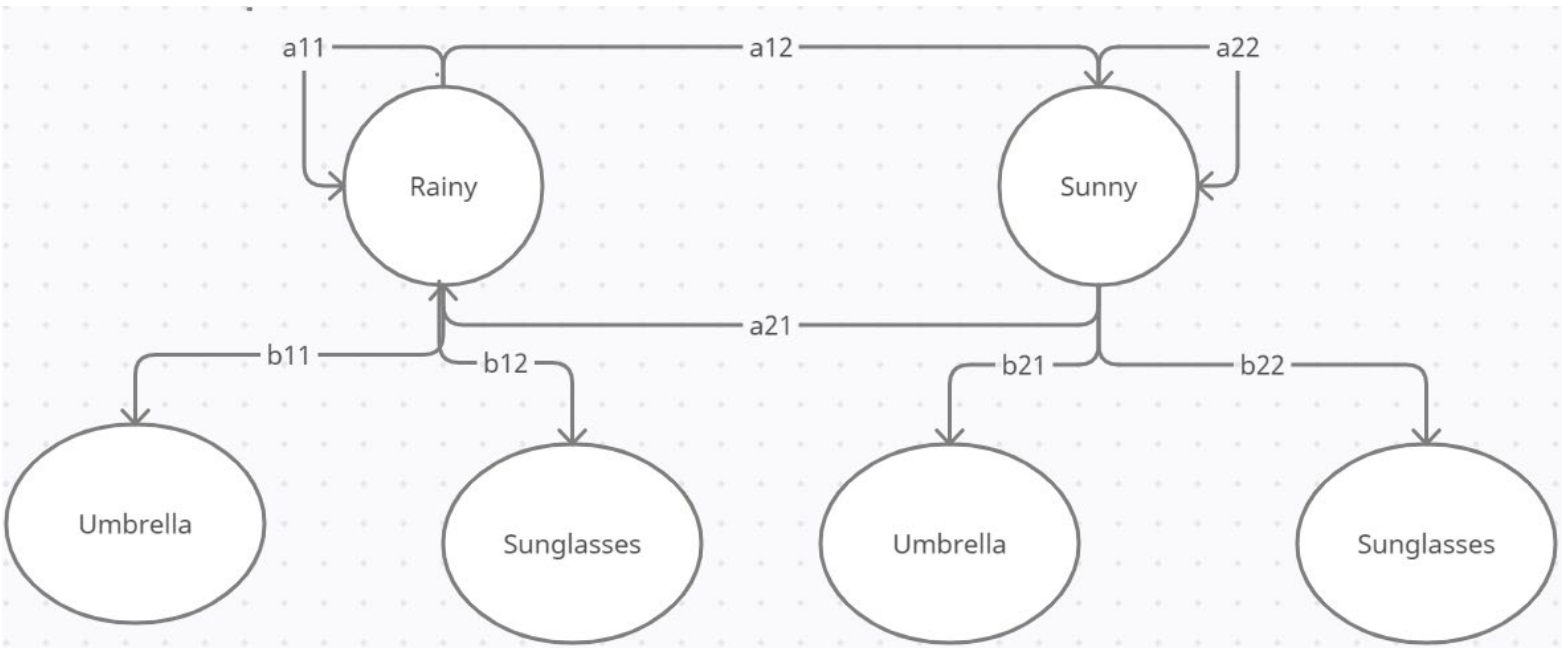
- **Non-proliferation goal:** monitor manufacturing and testing processes that might present a proliferation risk.

- **Problem:** data collected from monitoring systems does not yield direct knowledge of the activity underway.

Proliferation Monitoring Challenges

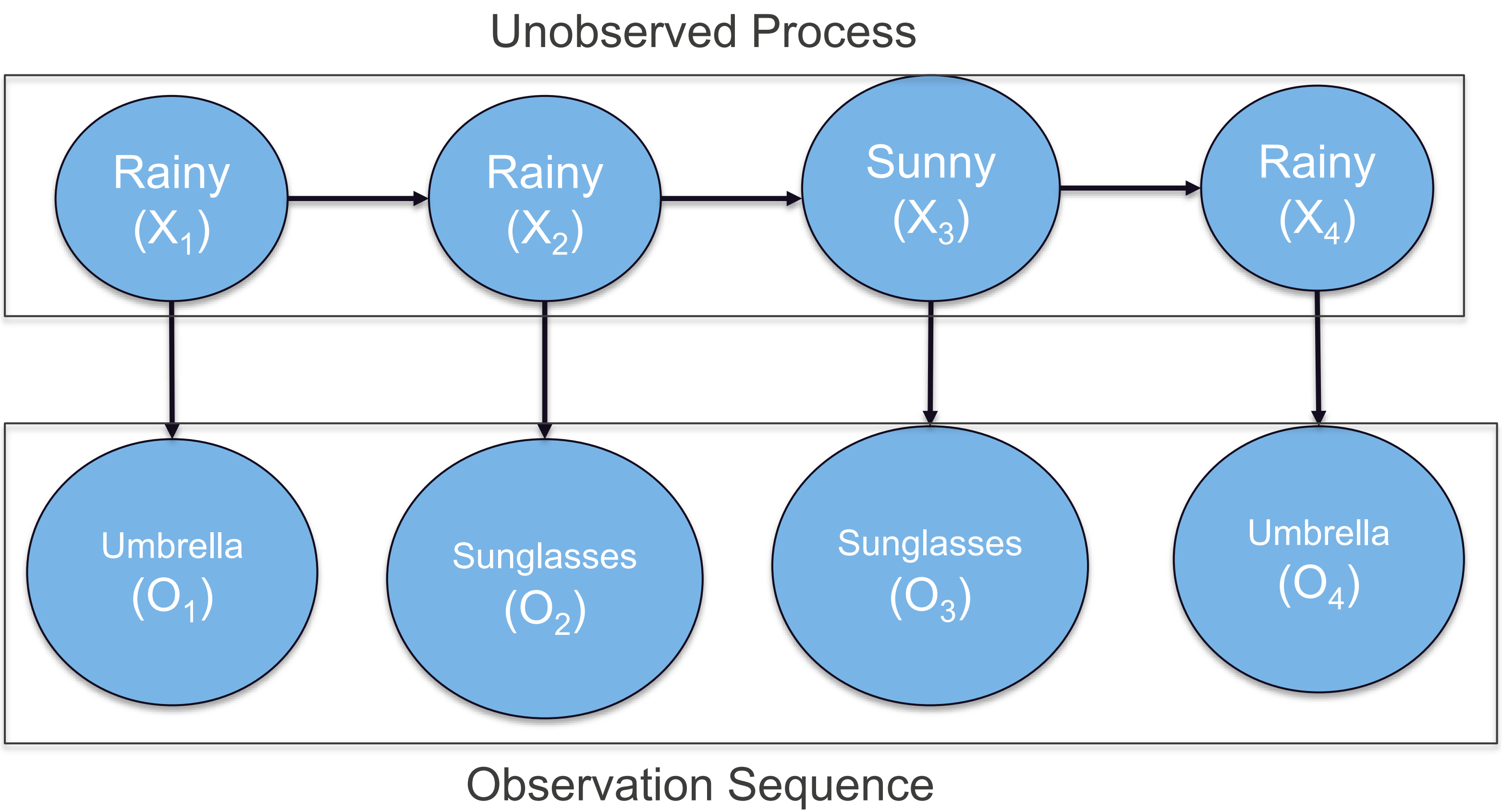
- **Goal:** develop a statistical model that combines
  - data observed from the process
  - domain knowledge about the process
- This model should describe
  - process of interest (unobserved)
  - process data (observed)
  - relationship between the process and the data

Simple HMM Example



Simple Weather HMM

HMM Data Stream



HMM Parameters

- **Initial State Probabilities**: $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)$, $\pi_i = P(X_1 = i)$

- **Observation Probabilities**: $B = \begin{pmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{N1} & \cdots & b_{Np} \end{pmatrix}$, $b_{ij} = P(O_t = j | X_t = i)$,

  $i = 1, \dots, N, j = 1, \dots, p,$ and $t = 1, \dots, n.$

- **Transition Probabilities:** $A = \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix}$, $a_{ij} = P(X_t = j | X_{t-1} = i)$,

  $i, j = 1, \dots, N,$ and $t = 1, \dots, n.$

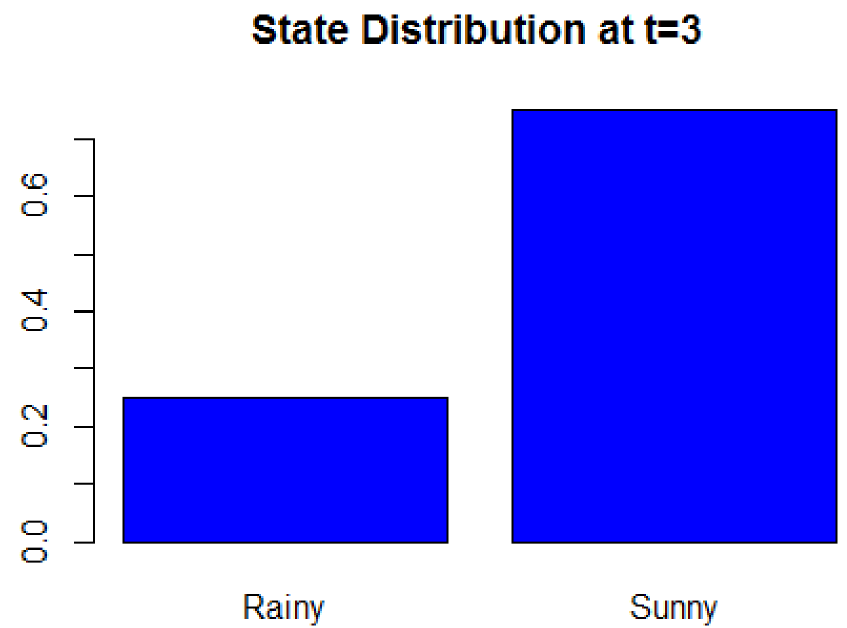
HMM Inference

- The HMM can be used to compute

$$\gamma_t(i) = \mathrm{P}(X_t = i | \boldsymbol{O}, \boldsymbol{\lambda}),$$

where $\boldsymbol{\lambda}$=(A, B, $\boldsymbol{\pi}$) and $\boldsymbol{O} = (O_1, \ldots, O_n)$.

- **Takeaway:** Infer most likely activity at any given time.



Dry Alluvium Geology (DAG) Test Case Study

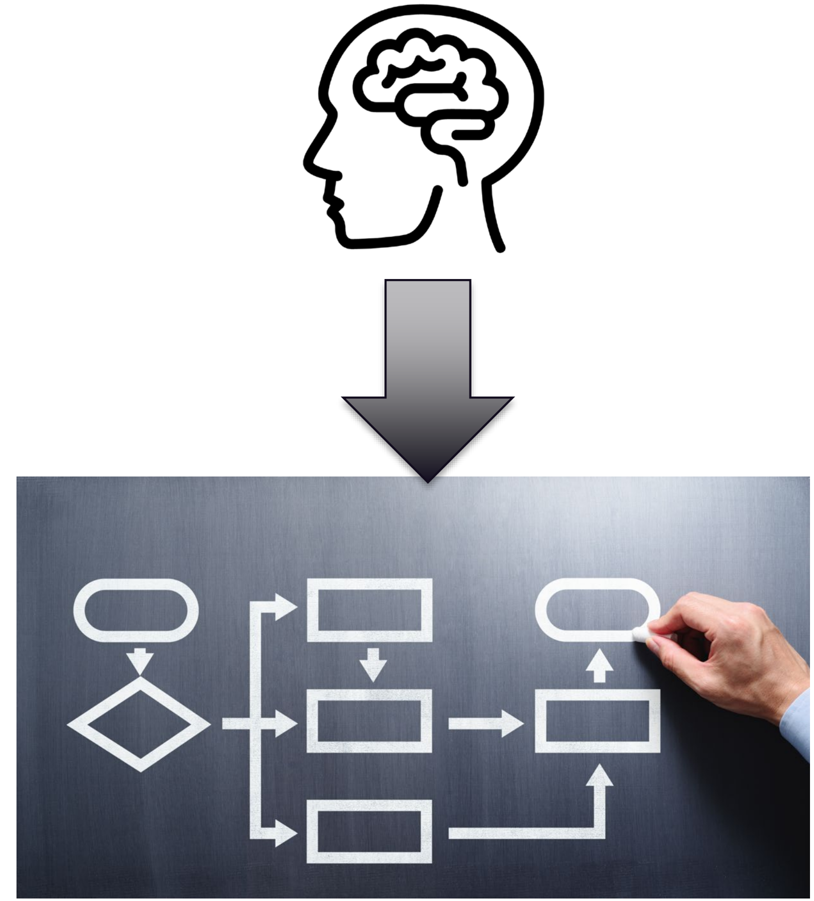
- Case study: Dry Alluvium Geology (DAG) test, an explosive test that was conducted at the Nevada National Security Site.

- Observation data: equipment (cranes, forklifts, etc.) in use at several evenly spaced time points.

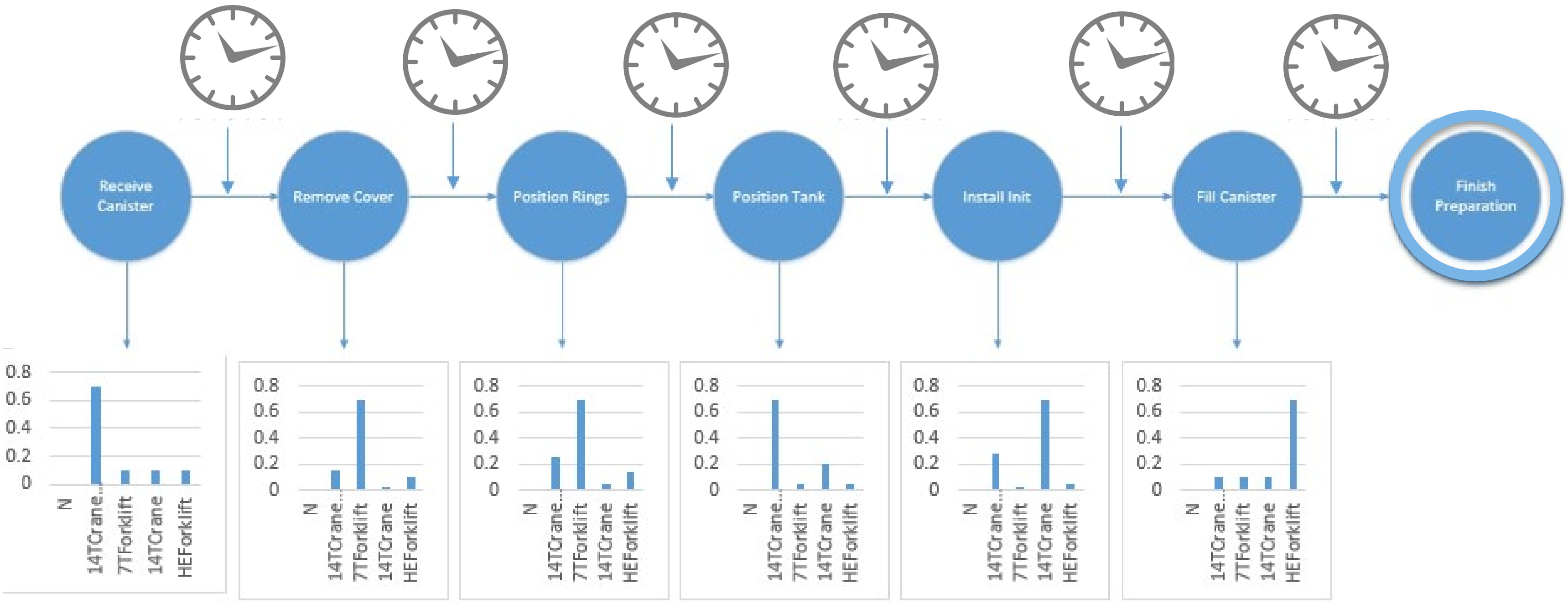
Domain Awareness: Parameterizing DAG HMM

- How do we incorporate expert knowledge to make our model domain aware?

- Discrete even simulator:
  - a process model built by experts
  - used to simulate DAG process runs
  - use runs to estimate observation probabilities and average activity completion times
  - transition probabilities can be derived from the average activity completion times

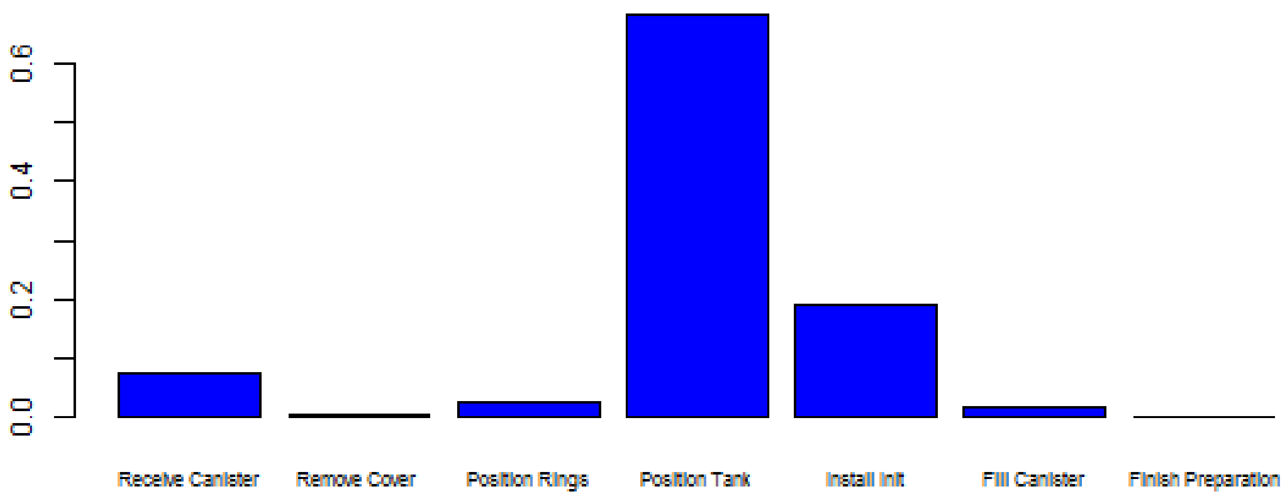


DAG HMM, cont.


Receive Canister
Remove Cover
Position Rings
Position Tank
Install Init
Fill Canister
Finish Preparation
0.8
0.6
0.4
0.2
0
N
14TCrane...
7TForklift
14TCrane
HEForklift

Determining the Most Likely Current Activity

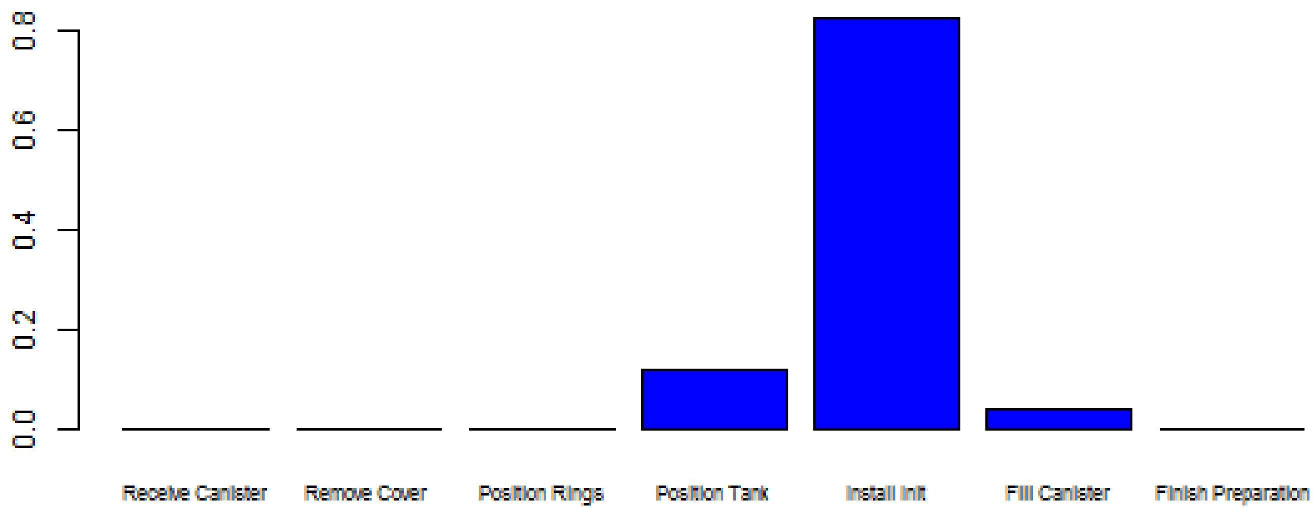
Initial observation sequence: (14T Crane/Old Glory, 7T Forklift, 7T Forklift, 14T Crane/Old Glory, 14T Crane/Old Glory)



Distribution over process activities after the final observation

Determining the Most Likely Current Activity (Cont.)

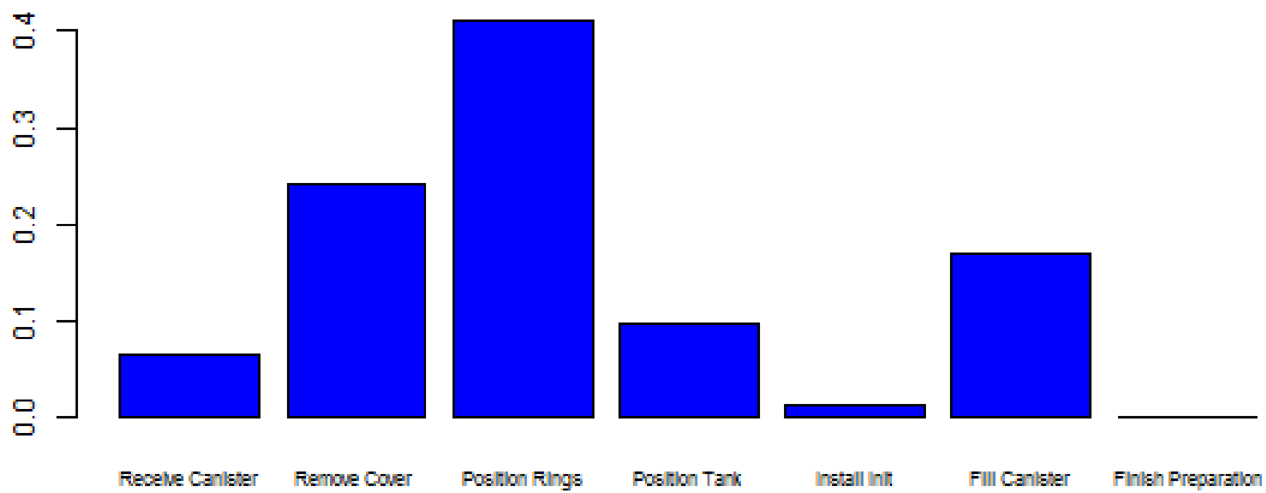
Replace last two observations with 14T Crane, which is indicative of the "install initiator" activity. Observation Sequence: (14T Crane/Old Glory, 7T Forklift, 7T Forklift, 14T Crane, 14T Crane)



Distribution over process activities after the final observation

Determining the Most Likely Current Activity (Cont.)

Longer observation sequence with more uncertainty: (14T Crane/Old Glory, 7T Forklift, 7T Forklift, 14T Crane/Old Glory, 14T Crane/Old Glory, 14T Crane/Old Glory, 7T Forklift, 7T Forklift)



Distribution over process activities after the final observation

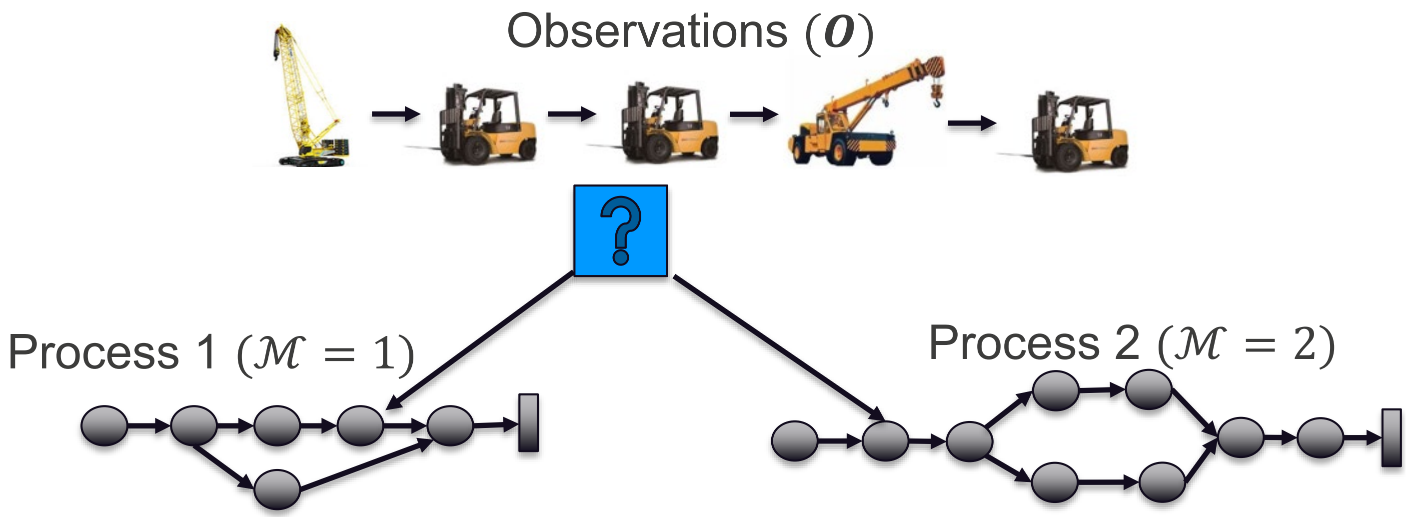
Other HMM capabilities

The HMM can also

- determine the most likely sequence of activities corresponding to a sequence of observations
- predict when the process started and when the process will end
- use observed data to update model parameters and quantify parameter uncertainty

Next Steps

Next step: Determine what process from a set of processes most likely generated the observed data.



We compute $P(\mathcal{M} = 1 | \boldsymbol{O})$ and $P(\mathcal{M} = 2 | \boldsymbol{O})$ and compare.

Thank you for your attention!