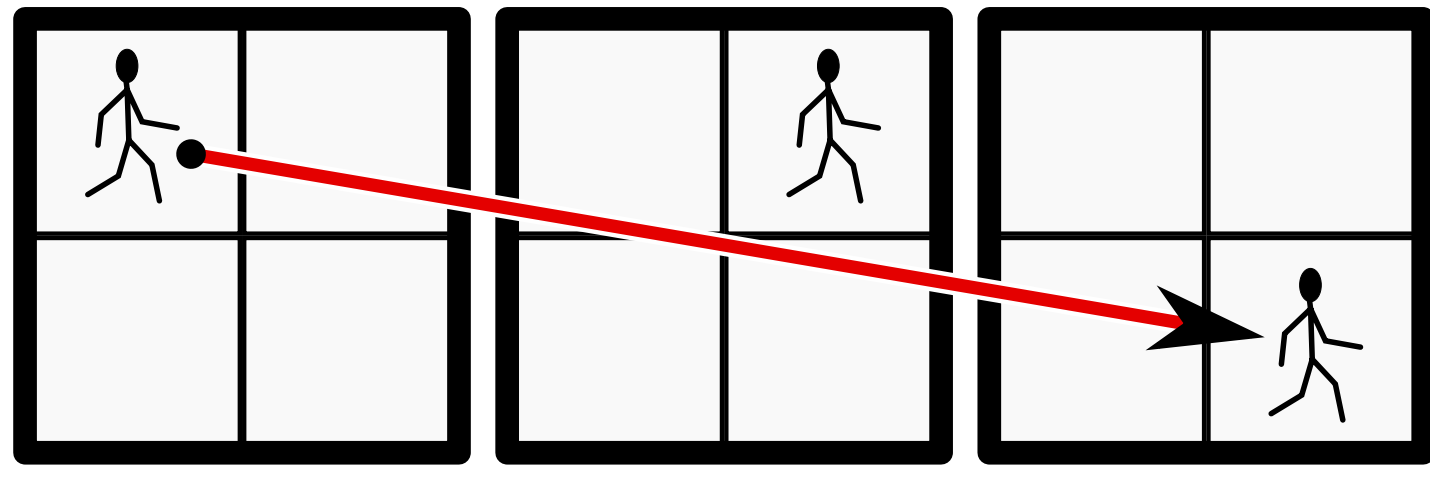
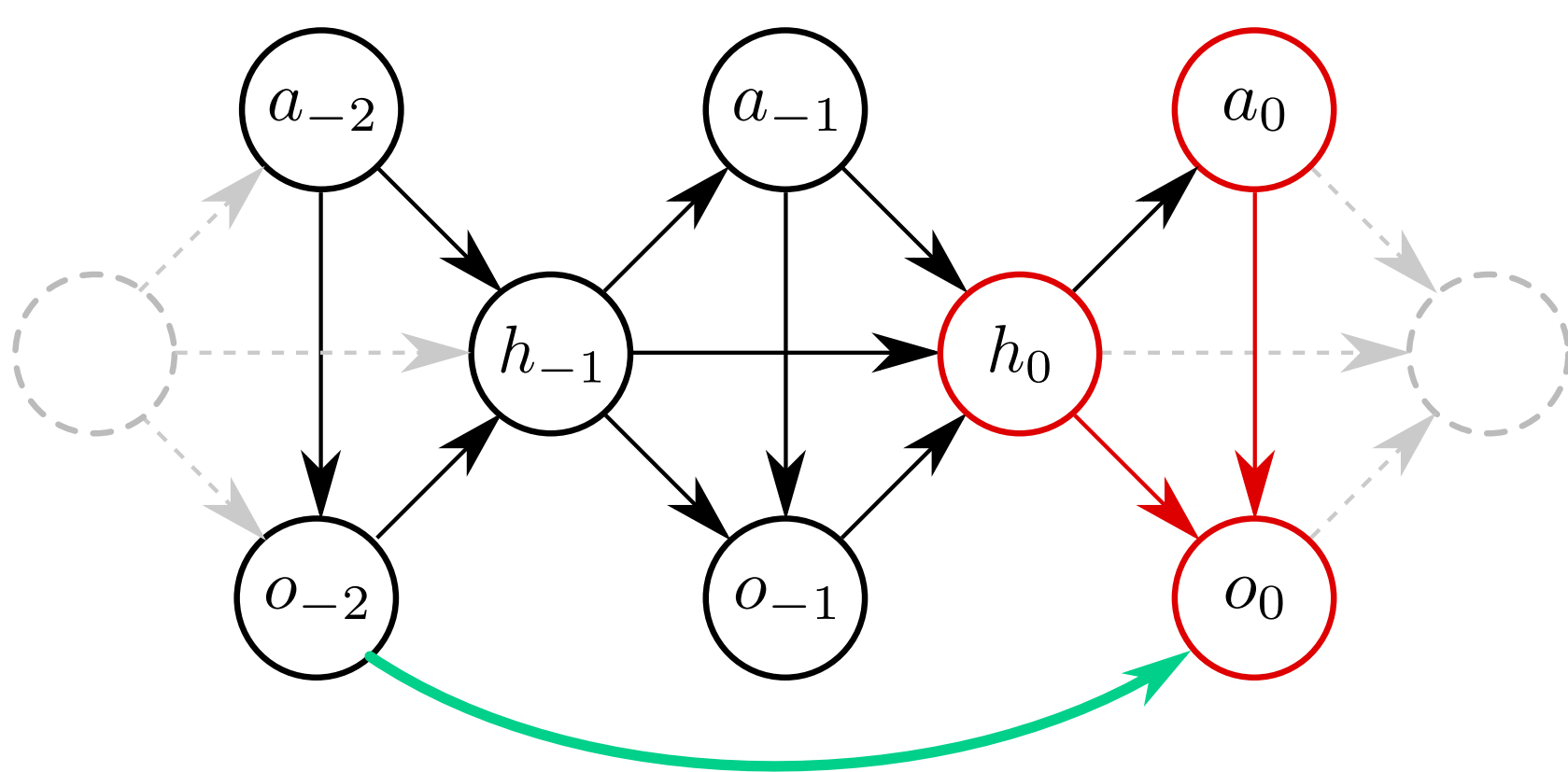


Events with Long-term Effects

A minimal example:
Rewards are "activated" and "collected".

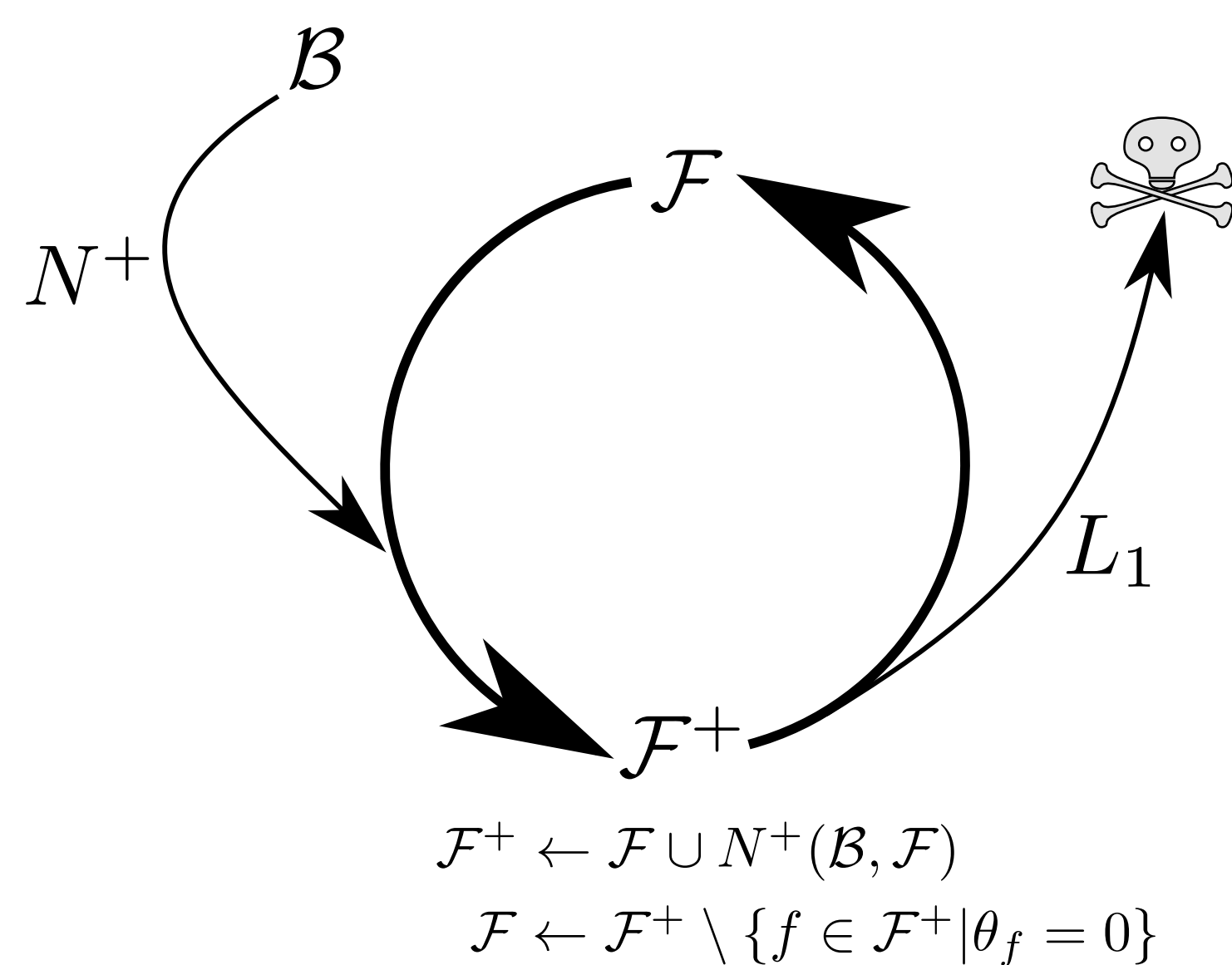


History as Markov state



Discovering the Features

Iterative Expansion and Pruning



The Algorithm

Input: $\mathcal{B}, N^+, D, \rho$
Output: \mathcal{F}, Θ

Initialize: $\mathcal{F} = \emptyset, \Theta = \emptyset$

repeat

 grow_feature_set($\mathcal{F}, \Theta, \mathcal{B}, N^+$)

 optimize_CRF($\mathcal{F}, \Theta, D, \rho$)

 shrink_feature_set(\mathcal{F}, Θ)

until data_likelihood(\mathcal{F}, Θ, D) does not change

grow_feature_set($\mathcal{F}, \Theta, \mathcal{B}, N^+$) {

 Initialize: $\mathcal{F}^+ = N^+(\mathcal{F})$

for all $f \in \mathcal{F}^+$ **do**

if $f \notin \mathcal{F}$ **then** $\theta_f = 0$ **endif**

end for

$\mathcal{F} = \mathcal{F} \cup \mathcal{F}^+$

}

shrink_feature_set(\mathcal{F}, Θ) {

for all $f \in \mathcal{F}$ **do**

if θ_f is 0 **then** $\mathcal{F} = \mathcal{F} \setminus f$ **endif**

end for

}

Temporally Extended Features

$$f = \begin{cases} 1 & \text{if } o_{-2} = \text{red} \wedge o_0 = \text{red} \wedge r_0 = \text{smiley} \\ 0 & \text{else} \end{cases}$$

The selected features (37 out of 64000)

Reward probability

$\checkmark(0)$ -11.91
 $\times(0)$ 11.91

What location can be reached with what action.

$\uparrow(0) \wedge \text{red}(0)$ 14.84
 $\uparrow(0) \wedge \text{red}(0)$ 15.11
 $\downarrow(0) \wedge \text{red}(0)$ 15.43
 $\downarrow(0) \wedge \text{red}(0)$ 15.69
 $\leftarrow(0) \wedge \text{red}(0)$ 14.86
 $\leftarrow(0) \wedge \text{red}(0)$ 14.78
 $\rightarrow(0) \wedge \text{red}(0)$ 15.60
 $\rightarrow(0) \wedge \text{red}(0)$ 14.96

What location can be reached from what other location.

$\text{red}(0) \wedge \text{red}(-1)$ 12.95
 $\text{red}(0) \wedge \text{red}(-1)$ 7.15
 $\text{red}(0) \wedge \text{red}(-1)$ 7.03
 $\text{red}(0) \wedge \text{red}(-1)$ 16.02
 $\text{red}(0) \wedge \text{red}(-1)$ 8.18
 $\text{red}(0) \wedge \text{red}(-1)$ 7.73
 $\text{red}(0) \wedge \text{red}(-1)$ 15.60
 $\text{red}(0) \wedge \text{red}(-1)$ 7.97
 $\text{red}(0) \wedge \text{red}(-1)$ 7.37
 $\text{red}(0) \wedge \text{red}(-1)$ 14.36
 $\text{red}(0) \wedge \text{red}(-1)$ 9.34
 $\text{red}(0) \wedge \text{red}(-1)$ 6.28

Correlation between rewards and locations.

$\checkmark(0) \wedge \text{red}(0)$ 13.50
 $\times(0) \wedge \text{red}(-1)$ 8.22
 $\times(0) \wedge \text{red}(-1)$ 6.59
 $\checkmark(0) \wedge \text{red}(-2)$ 7.71
 $\times(0) \wedge \text{red}(-2)$ -7.71

Other correlations.

$\text{red}(0) \wedge \text{red}(-2)$ -6.67
 $\text{red}(0) \wedge \text{red}(-2)$ -3.37
 $\text{red}(0) \wedge \text{red}(-2)$ -2.37
 $\text{red}(0) \wedge \text{red}(-2)$ -0.95
 $\text{red}(0) \wedge \leftarrow(-1)$ -3.23
 $\text{red}(0) \wedge \uparrow(-1)$ 2.78
 $\text{red}(0) \wedge \leftarrow(-1)$ 2.56
 $\text{red}(0) \wedge \rightarrow(-1)$ 1.31
 $\times(0) \wedge \text{red}(-1)$ -1.13
 $\checkmark(0) \wedge \text{red}(-1)$ 1.13

Building a Model

Conditional Random Fields

$$p(r_0, o_0 | a_0, h_0) = \frac{1}{Z(a_0, h_0)} \exp \sum_{f \in \mathcal{F}} \theta_f f(r_0, o_0, a_0, h_0)$$

$$Z(a_0, h_0) = \sum_{r_0, o_0} \exp \sum_{f \in \mathcal{F}} \theta_f f(r_0, o_0, a_0, h_0)$$

Learning Performance

