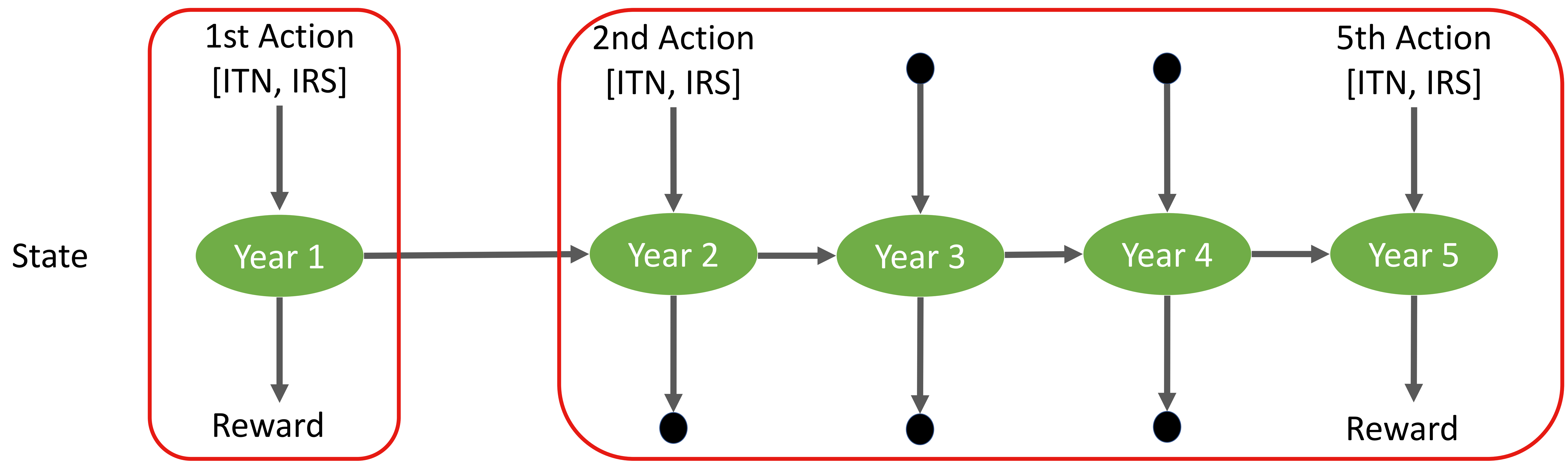
Thompson Sampling for Malaria Policy Interventions

Xiaolan Jiang
National Institute of Informatics
The Graduate University for Advanced Studies, Japan



Multi Armed Bandit

- Context: No
- Method: Thompson Sampling
- Use Beta function beta(a,b) to estimate the probability being the optimal action
- Initiliaze
  > a = 1
  > b = 3
- Update beta(a, b)
  > a = a + reward/150
  > b = b + 1 – reward/150

Contextual Multi Armed Bandit

- Context: Related to last year's (old) action
  > Each action has its own context
  > X = [abs(action^{ITN} - old^{ITN}), abs(action^{IRS} - old^{IRS}), abs(action^{ITN} - action^{IRS})]
- Method: Thompson Sampling
- Probability of Action j being the optimal: $(1+\exp(-\mathbf{w}^\top \mathbf{x}_j))^{-1}$
- How to update W: Alg.3 from paper: "An Empirical Evaluation of Thompson Sampling" by Olivier Chapelle, Lihong Li.

Require: Regularization parameter $\lambda > 0$.
$m_i = 0,\ q_i = \lambda$. {Each weight w_i has an independent prior $\mathcal{N}(m_i, q_i^{-1})$}
for $t = 1, \ldots, T$ **do**
Get a new batch of training data $(\mathbf{x}_j, y_j),\ \ j = 1, \ldots, n$.
Find $\mathbf{w}$ as the minimizer of: $\frac{1}{2}\sum_{i=1}^{d} q_i(w_i - m_i)^2 + \sum_{j=1}^{n} \log(1+\exp(-y_j \mathbf{w}^\top \mathbf{x}_j))$.
$m_i = w_i$
$q_i = q_i + \sum_{j=1}^{n} x_{ij}^2 p_j(1-p_j),\ \ p_j = (1+\exp(-\mathbf{w}^\top \mathbf{x}_j))^{-1}$ {Laplace approximation}
end for

Email : federerjiang@gmail.com