Modeling Adaptive Sampling Problems in Graphical Models using Markov Decision Process

Mathieu BONNEAU Nathalie PEYRARD Régis SABBADIN

INRA-MIA Toulouse
E-Mail: {mbonneau,peyrard,sabbadin}@toulouse.inra.fr

ECCS, Lisbon, September 2010

Motivation

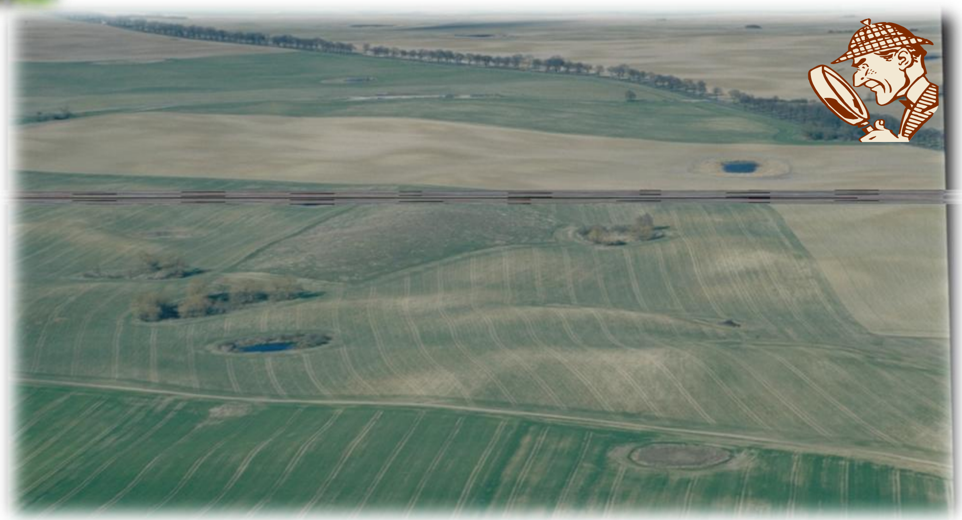
- Management/Control of a system are based on the whole map of the system:
  - Observation of the system is costly
  - Observations may be noisy

- **Problem:** Choose the observations which will be made to reconstruct the whole map of the system, taking sampling cost into account

Motivation: site-specific weed management

- **Context:** Traditionally herbicides are sprayed all over the field, whereas spraying can be limited to the infected area

=> Map of weeds populations



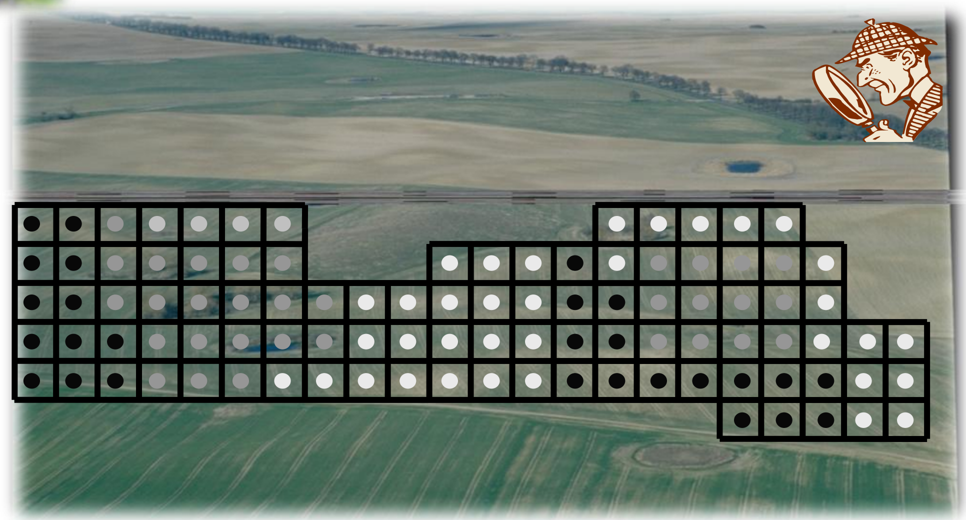
Problem: Fields are too large to be fully explored

=>Need to develop a sampling method

Motivation: site-specific weed management

- **Context:** Traditionally herbicides are sprayed all over the field, whereas spraying can be limited to the infected area

=> Map of weeds populations



Problem: Fields are too large to be fully explored

=>Need to develop a sampling method

Proposed method

A mathematical framework to define the adaptive sampling problem using graphical model

Optimal adaptive sampling problem ⇔ Quality optimization of sampling policy

☹

Proposed method

A mathematical framework to define the adaptive sampling problem using graphical model

Optimal adaptive sampling problem ⇔ ~~Quality optimization of sampling policy~~

PROBLEM TOO LARGE TO SOLVE EXACTLY

⇒Use simulation-based algorithm like reinforcement learning

⇒Model the adaptive sampling problem using Markov Decision Process (MDP)

Proposed method

☺

A mathematical framework to define the adaptive sampling problem using graphical model

⇓

Optimal adaptive sampling problem ⇔ Solving Markov Decision Process

PROBLEM TOO LARGE TO SOLVE EXACTLY

⇒Use simulation-based algorithm like reinforcement learning

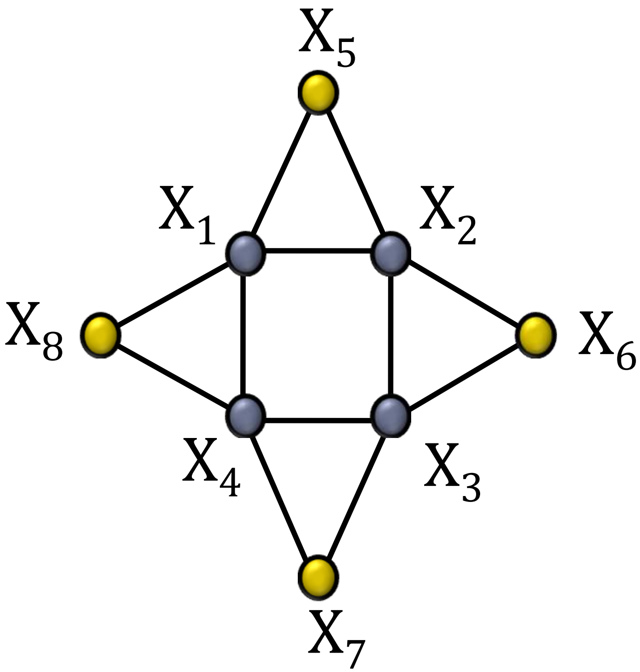
⇒Model the adaptive sampling problem using Markov Decision Process (MDP)

Related works

- Krause, *Phd thesis 2008:*
  - Adaptive sampling in Markov chain
  - Quality of policy based on entropy
  - Approximate solution using *greedy algorithm*

- Peyrard *et al*. *ECCS 2010:*
  - Adaptive sampling in Hidden Markov random field
  - Quality of sampling policy based on MAP
  - *Naive heuristics*

General Sampling Problem



$O = \{ 5 , 6 , 7 , 8 \}$

$R = \{ 1 , 2 , 3 , 4 \}$

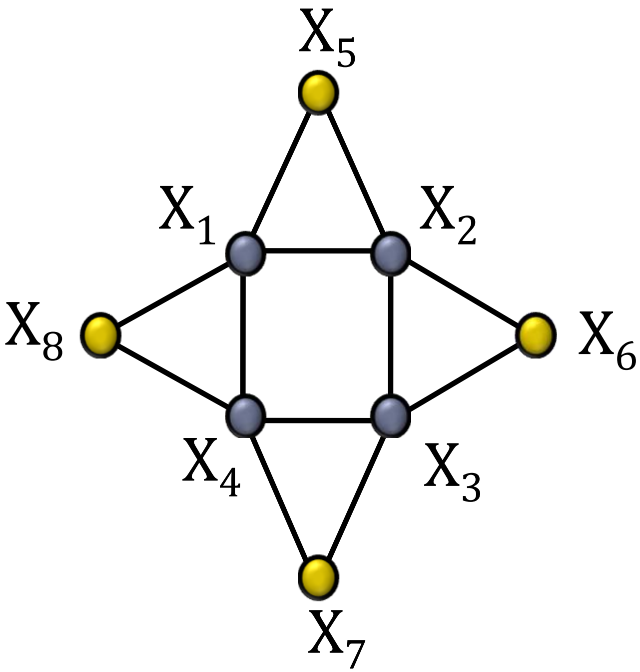
- Let $X=(X_1,\ldots,X_n)$ be a discrete random vector taking values into $\{ 0 , \ldots , K \}$

✓ **Goal :** Reconstruct the vector X_R

✓ **Difficulties :** Observations are available only on a subset of indices of O.

- $\{ 1 , \ldots , n \} = R \cup O$

General Sampling Problem



$O = \{5, 6, 7, 8\}$

$R = \{1, 2, 3, 4\}$

- Let $X = (X_1, \ldots, X_n)$ be a discrete random vector taking values into $\{0, \ldots, K\}$

✓ **<u>Goal :</u>** Reconstruct the vector X_R

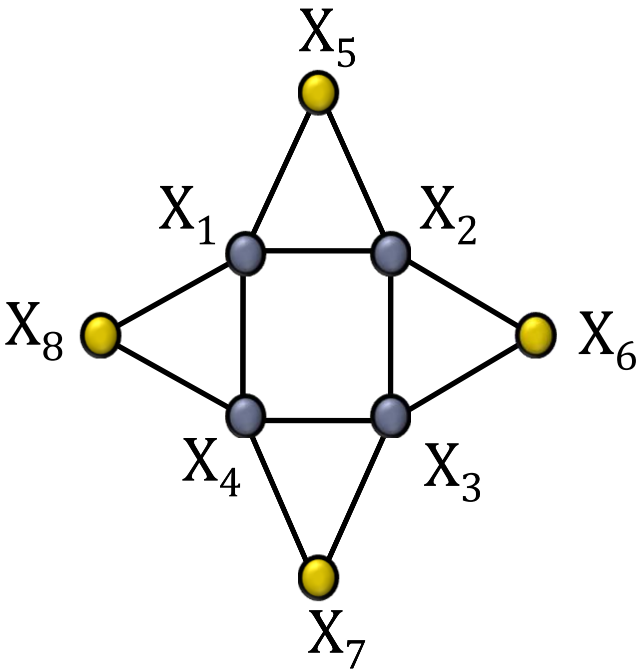
✓ **<u>Difficulties :</u>** Observations are available only on a subset of indices of O.

- $\{1, \ldots, n\} = R \cup O$

- $X \sim \mathbb{P}(. \mid \theta)$, a non-oriented graphical model

- We suppose that θ is known

General Sampling Problem



$O = \{ 5, 6, 7, 8 \}$

$R = \{ 1, 2, 3, 4 \}$

- Let $X = (X_1, \ldots, X_n)$ be a discrete random vector taking values into $\{ 0, \ldots, K \}$

✓ **Goal :** Reconstruct the vector X_R

✓ **Difficulties :** Observations are available only on a subset of indices of O.

- $\{ 1, \ldots, n \} = R \cup O$

- $X \sim \mathbb{P}(. | \, \theta)$, a non-oriented graphical model

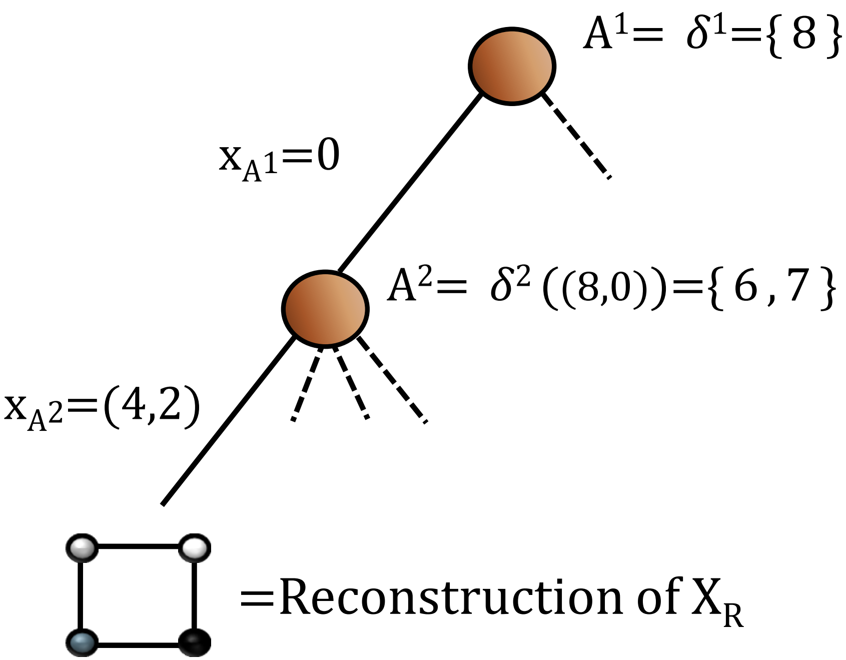
- We suppose that θ is known

⟹ Adaptively choose *sampling plans* $A^1, \ldots, A^H \subseteq O$ in order to reconstruct X_R

⟺ Choose a sampling policy

Sampling policy

Sampling policy of depth 2



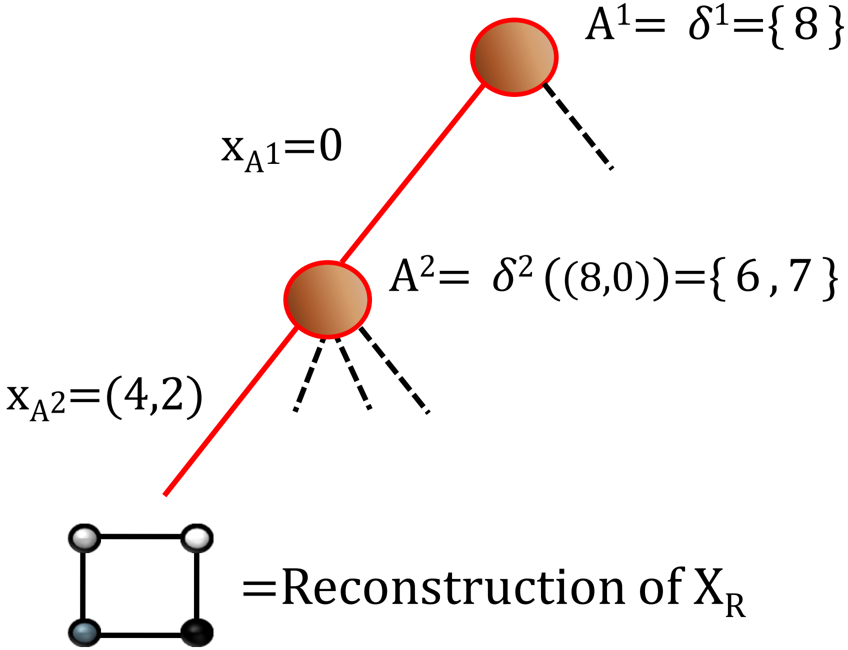
A sampling policy δ of depth H is a set of functions $(\delta^1, \ldots, \delta^H)$ such that:

$$A^1 = \delta^1$$

$$A^i = \delta^i((A^1, x_{A^1}), \ldots, (A^{i-1}, x_{A^{i-1}}))$$

Sampling policy

Sampling policy of depth 2



A sampling policy δ of depth H is a set of functions $(\delta^1,\ldots,\delta^H)$ such that:

$$A^1=\delta^1$$

$$A^i=\delta^i((A^1,x_{A^1}),\ldots,(A^{i-1},x_{A^{i-1}}))$$

A trajectory is a sequence of samples $\{(A^1,x_{A^1}),\ldots,(A^H,x_{A^H})\}$ issued from δ

τ_δ : the set of all possible trajectories issued from δ

Reconstruction of X_R and optimal sampling policy

- MAP reconstruction of X_R:

$$\mathbf{x}_\mathbf{R}^* = \operatorname*{argmax}_{\mathbf{x}_\mathbf{R}} \mathbb{P}(\mathbf{x}_\mathbf{R} \mid \mathbf{x}_{\mathbf{A}^1,\ldots,\mathbf{A}^H}, \theta)$$

- Trajectory quality:

$$V^{\mathrm{MAP}}((A^1, x_{A^1}),\ldots,(A^H, x_{A^H})) = \mathbb{P}(x_R^* \mid x_{A^1},\ldots,x_{A^H},\theta) - \sum_{i=1}^{H} c(A^i)$$

- Quality of a sampling policy:

$$\mathrm{Q}(\delta) = \sum_{(\mathbf{A},\mathbf{x}_\mathbf{A})\in\tau_\delta} \mathbb{P}(\mathbf{x}_\mathbf{A} \mid \theta)\mathrm{V}^{\mathrm{MAP}}((\mathbf{A},\mathbf{x}_\mathbf{A}))$$

- Optimal sampling policy:

$$\delta^* = \operatorname*{argmax}_{\delta} \mathrm{Q}(\delta)$$

Finite horizon Markov Decision Process

Definition

A MDP is defined as a 5-tuple $< S , D , T , \mathbf{P} , R >$:

- $T = \{ 1 , \ldots , H \}$. Finite set of decision steps
- S^t. Finite set of possible states of the system at time t
- D^t. Finite set of allowed decisions (or actions) at time t
- $\mathbf{P}_{d^t}(s^{t+1} | s^t)$. Transition probabilities
- $r^t(s^t, d^t)$. Immediate reward function at time t

Finite horizon Markov Decision Process

Definition

A MDP is defined as a 5-tuple < S , D , T , **P** , R > :

- $T = \{1, \ldots, H\}$. Finite set of decision steps
- S^t. Finite set of possible states of the system at time t
- D^t. Finite set of allowed decisions (or actions) at time t
- $\mathbf{P}_{d^t}(s^{t+1} | s^t)$. Transition probabilities
- $r^t(s^t, d^t)$. Immediate reward function at time t

- Policy : $\delta=(\delta^t)_{t=1\ldots H}$, where $\delta^t: \quad S^t \longrightarrow D^t$

- Criterion : $$V^\delta(s^1) = \mathrm{E}\left[\sum_{t=1}^{H} r^t(s^t, d^t) + r^{H+1}(s^{H+1}) \middle| \delta\right]$$

Finite horizon Markov Decision Process

Definition

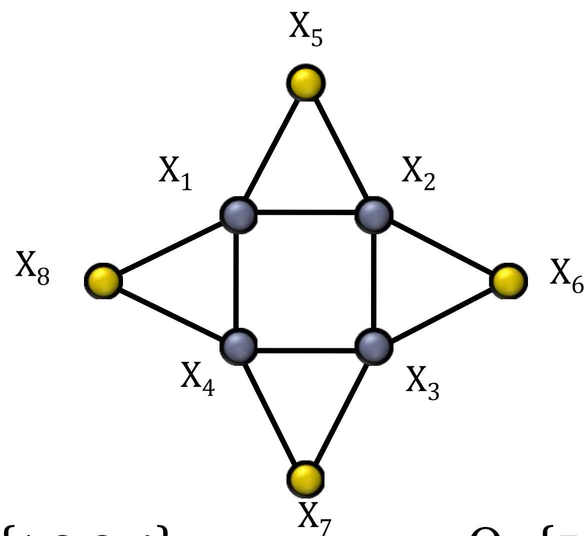
A MDP is defined as a 5-tuple < S , D , T , **P** , R > :

- T = { 1 , … , H } . Finite set of decision steps
- S^t. Finite set of possible states of the system at time t
- D^t. Finite set of allowed decisions (or actions) at time t
- $\mathbf{P}_{d^t}(s^{t+1} | s^t)$. Transition probabilities
- $r^t(s^t, d^t)$. Immediate reward function at time t

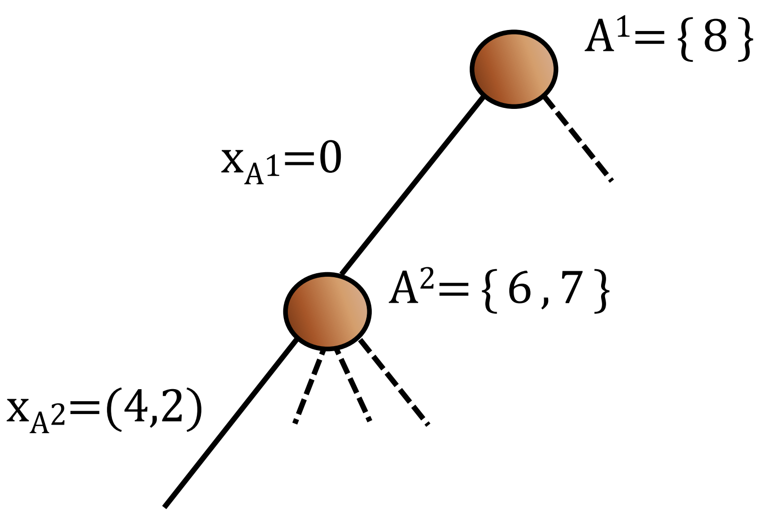
- Policy : $\delta=(\delta^t)_{t=1\ldots H}$, where $\delta^t: \quad S^t \longrightarrow D^t$
- Criterion : $$V^{\delta}(s^1) = \mathrm{E}\left[\sum_{t=1}^{H} r^t(s^t, d^t) + r^{H+1}(s^{H+1}) \middle| \delta\right]$$

Find an optimal policy δ^*, such that $\quad V^{\delta^*}(s^1) \geq V^{\delta}(s^1) \quad \forall \delta$

State and decision spaces



$R=\{1,2,3,4\}$ $O=\{5,6,7,8\}$



$s^1 = \varnothing$

$d^1=A^1=\{8\}$

$$s^2 = \begin{pmatrix} -1 \\ -1 \\ -1 \\ 0 \end{pmatrix}$$

$d^2=A^2=\{6,7\}$

$$s^3 = \begin{pmatrix} -1 & -1 \\ -1 & 4 \\ -1 & 2 \\ 0 & -1 \end{pmatrix}$$

Transition probabilities and reward function

- At time $t \in \{1, \ldots, H\}$, transition probabilities and reward functions:

$$\mathbf{P}_{d^t}(s^{t+1} \mid s^t) = \mathbb{P}\big(x_{A^t} \mid (A^1, x_{A^1}), \ldots, (A^{t-1}, x_{A^{t-1}}), \theta\big)$$

$$r^t(s^t, d^t) = -c(A^t)$$

- At time t=H+1, no decision is available but a global reward is attributed:

$$r^{H+1}(s^{H+1}) = V^{\mathrm{MAP}}\big((A^1, x_{A^1}), \ldots, (A^H, x_{A^H})\big) + \sum_{i=1}^{H} c(A^i)$$

Conclusion

- Solve:

$$\delta^* = \operatorname*{argmax}_{\delta} Q(\delta)$$

is equivalent to fine the optimal policy of our PDM

➡ Use simulation-based algorithm to solve the adaptive sampling problem in graphical model

Perspectives

- Design reinforcement learning algorithm using simulation method for graphical models
- Application to weeds mapping

THANK YOU!