Decision-theoretic Optimal Sampling in Hidden Markov Random Fields

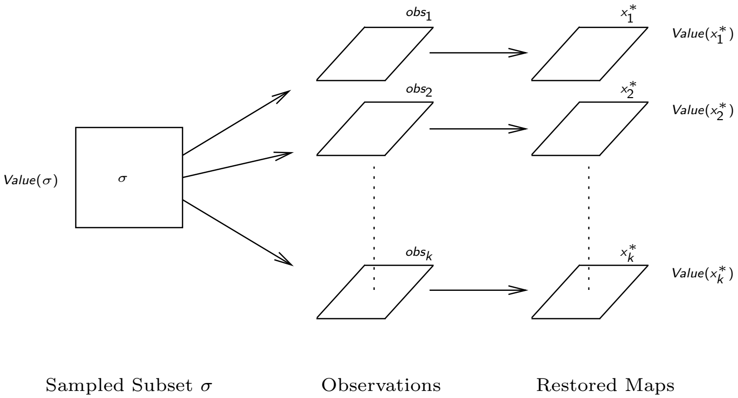
Nathalie PEYRARD, Régis SABBADIN and Usman F. NIAZ

MSTGA - Angers - June 2-3, 2010

- Motivation : Sampling optimisation for
  - Reconstruction of maps for *spatial systems*...
    - Invasive or protected species
    - Weeds or pests in a crop
  - ... or correlated ill-observed variables
    - Animal epidemics in farm networks
    - Pollution in water networks
    - Fault diagnosis in computer networks
- Objectives
  - Propose a HMRF-based model for such sampling problem
  - Understand algorithmic difficulties of such problems
  - Propose exact and approximate solution algorithms

- Principle : Find a subset of sites (variables) to sample
  - leading to the construction of a map of high quality
  - trading-off sampling cost and quality of the restored map
- Different types of sampling :
  - **Static Sampling**
  - Adaptive (or sequential) Sampling

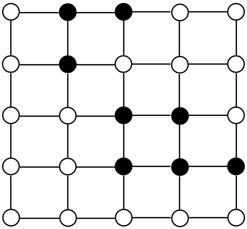
Static Sampling process



Map restoration in HMRF

- The state of each site is represented by a (random) variable X_i taking values $x_i \in \{0, ..., k-1\}$
- The spatial structure is represented by a graph $G = (V, E)$ showing local dependencies between variables
- Example : Presence/absence map

$$x_i = \begin{cases} 1 & \text{Presence of the phenomenon} \\ 0 & \text{Absence of the phenomenon} \end{cases}$$



Markov Random Fields

- $X = \{X_1, \ldots, X_n\}$: random variables taking values in $\{0, ..., k-1\}$
- P : a probability distribution over $\mathcal{X} = \{0, ..., k-1\}^n$
- $G = (V, E)$ with $|V| = n$ and $\mathcal{C}$ be the set of *cliques* of graph G
- $\Psi = \{\psi_c\}_{c \in \mathcal{C}}, \psi_c(x_c) > 0, \forall x_c$, a set of positive functions

Definition

P is a *Markov Random Field* defined by $(G, \mathcal{X}, \Psi)$ iff

$$P(X = x) = P_\Psi(x) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(x_c), \forall x \in \mathcal{X},$$

where Z is a normalising constant ensuring that P sums to 1.

Hidden Markov Random Field

- Variables X_i are not observed directly
- $Y = \{Y_1, \ldots, Y_n\}$ taking values in $\mathcal{Y} = \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_n$ are observed instead
- Y_i is conditionally independent of Y_j given X :

$$P(Y = y|X = x) = P_\Theta(y|x) = \prod_{i \in V} P_{\theta_i}(y_i|x_i), \forall x, \forall y.$$

where $\Theta = \{\theta_1, \ldots, \theta_n\}$ is a set of parameters.

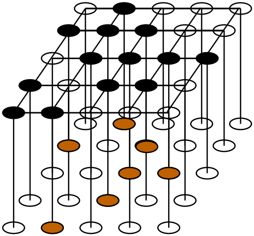
Definition

A *Hidden Markov Random Field (HMRF)* is defined by the tuple $(G, \mathcal{X}, \mathcal{Y}, \Psi, \Theta)$.

Hidden Markov Random Field

- Example

$y_i \backslash x_i$	0	1
0	1	$1-\theta$
1	0	θ



Hidden variables restoration in HMRF

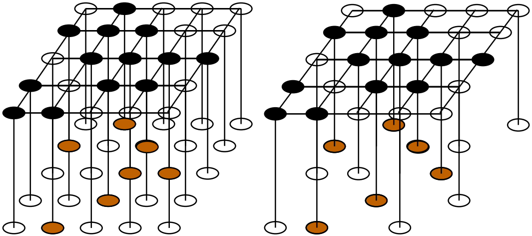
- Input :
  - $P_\Psi(x)$: MRF hidden variables distribution
  - $P_\Theta(y/x)$: HMRF observation function
  - *Observations y*
- Output :
  - Hidden variables posterior distribution :

$$P_{\Psi,\Theta}(x/y) = \frac{P_\Theta(y/x)P_\Psi(x)}{P_{\Psi,\Theta}(y)}$$

    where $P_{\Psi,\Theta}(y) = \sum_x P(y/x)P(x)$
  - Hidden variables restoration x^*

How to sample under limited resources ?

- Observations require to consume limited resources (time, people, etc.)
- A *sample* is a subset $\sigma \subseteq V$ of sites which will be actually observed in order to restore the hidden field x
- The result of a sample σ is a a *sample output* y_σ
- How to choose a *sample* $\sigma \subseteq V$ to observe ?



How to sample under limited resources ?

- Observations require to consume limited resources (time, people, etc.)
- A *sample* is a subset $\sigma \subseteq V$ of sites which will be actually observed in order to restore the hidden field x
- The result of a sample σ is a a *sample output* y_σ
- How to choose a *sample* $\sigma \subseteq V$ to observe ?

Definition (Posterior probability)

- HMRF $(G = (V, E), \mathcal{X}, \mathcal{Y}, \Psi, \Theta)$.
- $\sigma \subseteq V$ is a sample and y_σ a sample outpout
- The posterior probability conditionned on sample output y_σ is defined as :

$$P_{\Psi,\Theta}(x|y_\sigma) \propto P_\Theta(y_\sigma|x)P_\Psi(x), \forall x \in \mathcal{X}.$$

Sample quality definition

- How to define $U_{\psi,\Theta}(\sigma)$, the quality of sample σ ?
- Choice based on $P_{\Psi,\Theta}(x|y_\sigma)$ and on sample cost $\gamma(\sigma)$
- y_σ unkown when σ chosen $\Rightarrow$ expectation criterion

Definition (Decision-theoretic sample quality measure)

$$U_{\Psi,\Theta}(\sigma) = -\gamma(\sigma) + \sum_{y_\sigma} P_{\Psi,\Theta}(y_\sigma) V_{\Psi,\Theta}(\sigma, y_\sigma),$$

- $V_{\Psi,\Theta}$ measures how informative $P_{\Psi,\Theta}(x|y_\sigma)$ is
- $\gamma(\sigma)$ measures the "cost" of σ, and its unit is homogeneous to $V_{\Psi,\Theta}(\sigma, y_\sigma)$

Sample quality definition

- Computing $V_{\Psi,\Theta}(\sigma, y_\sigma)$ ressembles a problem of image restoration
- Different possible choices : MPM, **MPE** (MAP), Entropy...

Definition (Most Probable Explanation (MPE) criterion)

- Restoration quality :

$$V^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma) = \max_{x \in \mathcal{X}} P_{\Psi,\Theta}(x|y_\sigma), \forall y_\sigma$$

- Hidden variables restoration :

$$x^{MPE}(y_\sigma) = \arg\max_{x \in \mathcal{X}} P_{\Psi,\Theta}(x|y_\sigma), \forall y_\sigma$$

Decision-theoretic sample optimisation problem

Definition (Decision-theoretic sample optimisation problem)

- HMRF $(G = (V, E), \mathcal{X}, \mathcal{Y}, \Psi, \Theta)$
- integer $K \geq 0$, a set $W \subseteq V$ of sites available for sample and costs $\gamma_i \geq 0, \forall i \in W$

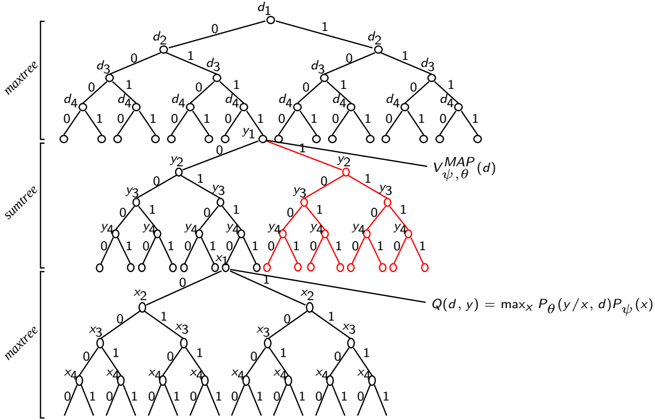
HMRF sample optimisation problem $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ defined as :

$$\text{Find } \sigma^* = \arg \max_{\sigma \subseteq W, |\sigma| \leq K} U^{MPE}_{\Psi,\Theta}(\sigma).$$

- How difficult is this sample optimisation problem ?
  - $\Rightarrow$ Computational complexity results
  - $\Rightarrow$ Polynomial approximations
- Practical solution algorithms
  - $\Rightarrow$ Exact variable elimination algorithm
  - $\Rightarrow$ Approximate (exponential) “greedy” algorithm
  - $\Rightarrow$ Aprroximate “Belief propagation” algorithm

Optimal sample exact computation

$$\sigma^* = \arg\max_{\sigma} -\gamma(\sigma) + \sum_{y_\sigma} \max_{x} P_{\Psi,\Theta}(y_\sigma|x)P_\Psi(x),$$



Computational complexity results

Several cases depending on :

- Whether the sample size K is "small" (i.e. of unconditionally bounded size)
- Whether Z is known (or $\gamma(\sigma)$ constant)

Proposition (General case)

The HMRF sample optimisation problem $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ is PP*-hard, even when Z is known*

Proof :

- "Sum-tree" complexity dominates
- Reduction of the problem MAJSAT
- Complexity lower bound is likely to be strict

Computational complexity results

Proposition (Best case : Z known, K small)

The HMRF sample optimisation problem $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ is NPO-*complete when Z is known is K small*

Proof :

- "Max-tree" complexity dominates
- Hardness by reduction of the problem MAX 2-SAT

Polynomial approximations

Proposition (Negative result)

$\mathcal{O}(U_{\Psi,\Theta}^{MPE})$ *does not belong to* APX.

Proof :

- Meaning that there is no polynomial algorithm providing a constant ratio approximation
- Proof by reduction of the problem MAX 2-SAT

Polynomial approximations

Proposition (Positive result)

For a HMRF with **binary variables**, **null cost function** *(Z known) and* $P_{\theta_i}(y_i|x_i) > 0, \forall i, x_i, y_i$ *:*
$U^{MPE}_{\Psi,\Theta}(\sigma)$ *can be approximated in polynomial time within an* **instance-dependent** *approximation ratio* $\rho_{\Psi,\Theta,\sigma} = \kappa^{0.23}_{\Psi,\Theta,\sigma}$, *where*

$$\kappa_{\Psi,\Theta,\sigma} = \frac{\prod_{c\in\mathcal{C}} \max_{x_c} \psi_c(x_c) \prod_{i\in\sigma} \max_{x_i,y_i} P_{\theta_i}(y_i|x_i)}{\prod_{c\in\mathcal{C}} \min_{x_c} \psi_c(x_c) \prod_{i\in\sigma} \min_{x_i,y_i} P_{\theta_i}(y_i|x_i)}.$$

If furthermore the HMRF is pairwise, $\rho_{\Psi,\Theta,\sigma} = \kappa^{0.069}_{\Psi,\Theta,\sigma}$.

Proof :

- Adaptation of best known polynomial algorithms for MAX SAT and MAX 2-SAT

What does it mean in practice ?

- That the decision-theoretic optimal sampling in HMRF problem is hard to solve exactly... or very hard !
- Still we can explore several directions :
  1. Exact variable elimination algorithm
  2. Greedy variable elimination algorithm
  3. Belief Propagation (message passing) algorithm

Exact variable elimination algorithm

- Exhaustive enumeration of the sample $\times$ observation space (i.e. all pairs (σ, y_σ))
- Exact calculation of $U^{MPE}_{\Psi,\Theta}(\sigma)$
- Variable elimination used for
  - Exact calculation of $V^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma)$
  - Exact calculation of Z (once for a given problem)

Calculation of $V^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma)$

$$V^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma) = \frac{1}{Z} \max_{x_1,\ldots,x_n} \prod_{i \in V} P_{\theta_i}(y_i / x_i, \sigma) \prod_{c \in \mathcal{C}} \psi_c(x_c)$$

- Successive elimination of x_i through local variable elimination
- Find max on the potentials of each variable and project it on its neighbours
- Elimination order 1 to n

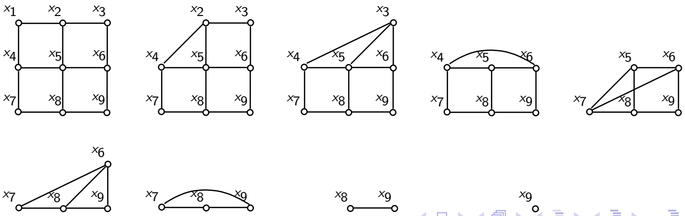
- For every x_k eliminated
  - Calculate

$$h_k = \max_{x_k} P_{\theta_k}(y_k/x_k, \sigma) \prod_{c)\in C, k\in c} \psi_c(x_c) h_{k-1}$$

  - Store h_k
  - New edges added between neighbours of x_k
- End when no variables left

$$V^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma) = \max_{x_n} P_\theta(y_n/x_n, \sigma) h_{n-1}$$

- Similar principle for computing Z



Greedy approximation algorithm

- The exact variable elimination algorithm requires to perform variable elimination for
  - All available samples σ (number exponential in K)
  - All possibles observations σ (number exponential in K)
- So, it needs to perform a task require exponential (in $|V|$) time many (exponential in $K\,!$) times

The idea of the greedy algorithm is to reduce (a bit) this complexity, by "greedily" exploring samples of increasing size :

- Complexity decreased by an exponential factor (but still exponential !)
- Unfortunately, no guarantee on the loss of quality by using the greedy algorithm (could have been if $U^{MPE}_{\Psi,\Theta}$ were *submodular*)
- Quite good in practice (see experiments) !

Belief propagation algorithm

Double approximation :

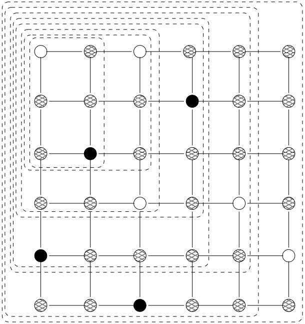
- Marginal probability distributions for each vertex computed *approximately* a priori, using a classical "Sum-Product" BP algorithm
- Sample choice made on the basis of these approximate distributions (most "uncertain" vertices chosen for sampling

Consequences :

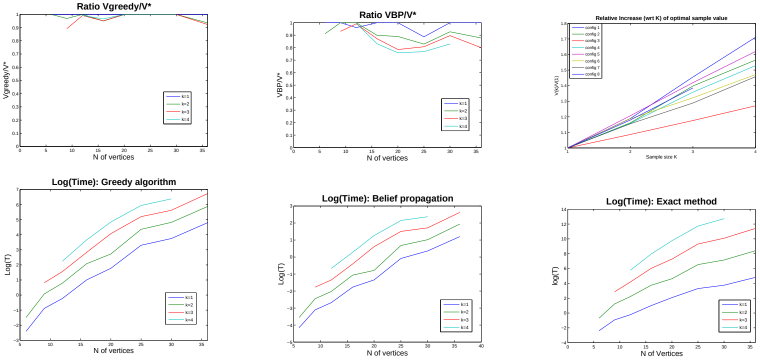
+ Sum Product BP algorithm run only once
+ No need to explore the sample output space
+ Complexity independent of sample size
+ Sum Product BP run time can be "tuned" (number of updates)
- Less efficient than the Greedy algorithm

Experiments

- Pairwise HMRF with $\mathcal{X}_i = \mathcal{Y}_i = \{0, 1\}$
- Singleton potentials : white ($\psi_i(1) = 0.3$), grey ($\psi_i(1) = 1$) black ($\psi_i(1) = 3$)
- Pair potentials : $\exp(1) = e$ for $(1, 1)$ and $(0, 0)$, 1 otherwise.



Results



- Top : Ratio of values, greedy (left), BP (center), relative increase of optimal value wrt sample size.(right).
- Bottom : Log-times of greedy, BP and exact algorithms.

Comments

- Exact method inapplicable for large problems (Couldn't solve the 6x6 problem with $k = 4$ within 9 days)
- Greedy algorithm considerably faster (x10 for $k = 2$, x100 for $k = 3$ and x500 for $k = 4$) and still close to optimal (for small problems).
- BP is the only applicable method for large problems (and time independent of K). Optimality ratio decreases to $< 80\%$.

Summary

- An approach for optimal sampling in structured problems under limited budget, within the framework of HMRF
- Criterion : tradeoff between expectatation (over outputs) of the *most likely* reconstruction and *sample cost*
- Results :
  1. Complexity/approximability results
  2. Exact and approximation algorithms (empirical validation)
- Originality :
  1. General HMRF structure with decisions
  2. Noisy observations
  3. Non-trivial quality measure
  4. Focus on sample size (rather than graph structure) to exhibit “easier” classes of problem

Related works

Several recent works have addressed the question of decision-theoretic observation selection in graphical models :

- [KG09] (i) Reliable observations (ii) simple problem structures (chain model, naive Bayes model or polytree). (iii) Complexity results and exact and approximate solution algorithms (iv) Rewards are *local*.
- [RSS06] (i) Noisy observations case (ii) Simple problem structures (hidden Markov chains, tree-shaped Bayesian networks) (iii) Easily computable local rewards
- [MS08] (i) General Bayesian networks (ii) Noisy observations, (iii) Specific and easy to compute non-local reward function.
- [PSS$^+$10] (i) Similar problem with application to invasive species map reconstruction (ii) MPM-based criterion (iii) static and adaptive case (iv) BP and heuristic algorithms.

Related works

Authors	Structure	Obs	Reward	Algo	Complexity
[KG09]	Simple	Reliable	Easy	Ex+App	Yes
[RSS06]	General	Noisy	Easy	Exact	No
[MS08]	General	Noisy	Easy	Exact	No
[PSS+10]	General	Noisy	MPM	App	No

Future work

Mathieu Bonneau's PhD thesis (with S. GABA, INRA-EA Avignon)

- Extending the framework to general stochastic graphical model
- Exploring simulation-based algorithms (Reinforcement Learning) for adaptive sampling problems with different criteria
- Comparing with Krigging-based approaches [BPS10]
- Weeds communities mapping application

M. Bonneau, N. Peyrard, and R. Sabbadin.
Echantillonnage spatial basé sur le krigeage pour la reconstruction de carte d'occurrence.
In *Proc. 17th Conf. Reconnaissance des Formes et IA (RFIA'2010)*, 2010.

A. Krause and C. Guestrin.
Optimal value of information in gaphical models.
Journal of Artificial Intelligence Research, 35 :557–591, 2009.

M. Munie and Y. Shoham.
Optimal testing of structured knowledge.
In *Proc. of 23rd AAAI Conference*, 2008.

N. Peyrard, R. Sabbadin, D. Spring, R. Mac Nally, and B. Brook.

Spatial sampling in hmrf mapping problems : Static and adaptive algorithms.
In *ECCS conference*, Lisbon, Portugal, 2010.

Y. Radovilsky, G. Shattah, and S. Shimony.
Efficient deterministic approximation algorithm for nonmyopic value of information in graphical models.
In *SMC conference*, Taipei, Taiwan, 2006.