Title

Spatial Sampling for Model Selection

Authors
Hajj Chehade, Nabil
Pottie, Greg

Publication Date
2007-10-10

Peer reviewed

CENS Center for Embedded Networked Sensing

Spatial Sampling for Model Selection

Nabil Hajj Chehade, **Greg Pottie**
EE, UCLA

Introduction

Environment Modeling Uncertainty

- **Wireless sensing systems are very useful for applications where we need to learn about environmental phenomena over spatial and temporal fields.**
- **Parametric models are widely used to represent these environmental fields and to answer questions and inferences regarding the phenomena.**
- **Choosing a model structure to represent the field involves a great deal of uncertainty.**
- **Often data are collected at random locations. We present methods for data collection to optimize the desired inferences.**
- **Often a single model *M* is used. If *M* does not characterize a phenomenon correctly, the inferences and predictions will not be accurate.**
- **It is better to start with *multiple* plausible models and select the model by collecting measurements at *informative* locations.**

Regression

- **The polynomial regression Model:**

$$t_i = \theta_0 + \theta_1 x_i + \cdots + \theta_n x_i^n + e_i, \qquad i = 1, \ldots, n$$

- **Vector form:**

$$\mathbf{t} = \mathbf{X}^T \grave{\mathbf{e}} + \mathbf{e}$$

- **We assume Gaussian noise, the ML estimate of θ is given by**

$$\hat{\mathbf{e}} = (\mathbf{X}\mathbf{C}_e^{-1}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{C}_e^{-1}\mathbf{t}$$

- **The estimation error covariance matrix is given by**

$$\mathbf{C}_{\tilde{e}} = (\mathbf{X}\mathbf{C}_e^{-1}\mathbf{X}^T)^{-1}$$

- **The error covariance matrix depends on the design matrix X and does not depend on the measurements.**

Reducing Uncertainty in Model Estimation

Problem Description: Optimal Sensor placement
Where should we collect measurements to optimally estimate the parameters of the regression model?

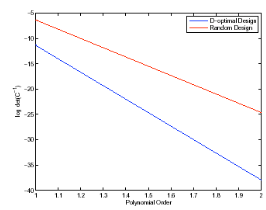
Assumptions: The model used *is the correct model*.
Gaussian noise ($\mathbf{C}_e = \mathbf{I}$).

Idea: Find the locations that result in "the smallest" error covariance matrix.

Technically:

$$\underset{\hat{\imath}}{Min} \ \det(\mathbf{X}\, diag(\hat{\imath})\mathbf{X}^T)^{-1}$$
$$s.t. \quad \mathbf{1}^T\hat{\imath} = 1$$
$$\hat{\imath} \geq \mathbf{0}$$

Algorithm: D-Designs
The minimization problem is convex and be solved with any convex optimization software.



Multiple model estimation:

$$\mathbf{t} = \mathbf{X}_t^T\grave{\mathbf{e}} + \mathbf{e}$$
$$\mathbf{l} = \mathbf{X}_l^T\grave{\mathbf{e}} + \mathbf{e}$$

We can setup three problems for this case:

1. $\underset{\hat{\imath}}{Min} \ \left(\det(\mathbf{X}_t\, \mathrm{diag}(\hat{\imath})\mathbf{X}_t^T)^{-1} + \det(\mathbf{X}_l\, \mathrm{diag}(\hat{\imath})\mathbf{X}_l^T)^{-1}\right)$
   $s.t. \quad \mathbf{1}^T\hat{\imath} = 1, \quad \hat{\imath} \geq \mathbf{0}$
2. $\underset{\hat{\imath}}{Min} \ \left(\det(\mathbf{X}_t\, \mathrm{diag}(\hat{\imath})\mathbf{X}_t^T)^{-1} \bullet \det(\mathbf{X}_l\, \mathrm{diag}(\hat{\imath})\mathbf{X}_l^T)^{-1}\right)$
   $s.t. \quad \mathbf{1}^T\hat{\imath} = 1, \quad \hat{\imath} \geq \mathbf{0}$
3. $\underset{\hat{\imath}}{Min} \ \left(\det(\mathbf{X}_t\, \mathrm{diag}(\hat{\imath})\mathbf{X}_t^T)^{-1}\right)$
   $s.t. \quad \left(\det(\mathbf{X}_l\, \mathrm{diag}(\hat{\imath})\mathbf{X}_l^T)^{-1}\right) \leq c, \quad \mathbf{1}^T\hat{\imath} = 1, \quad \hat{\imath} \geq \mathbf{0}$

Reducing Uncertainty in Model Selection

Problem Description: Optimal Sensor placement
Where should we collect measurements to optimally choose a model that represents the field?

Assumptions: *A set of* plausible models.
The set *contains* the correct model.
Gaussian noise.

Idea: Find the locations where the "difference" between the two models is the largest.

Derivation:
Two model case

$$M_1: t_i = \eta_1(x_i, \theta_1) + e_i, \qquad i = 1, \ldots, n$$
$$M_2: t_i = \eta_2(x_i, \theta_2) + e_i, \qquad i = 1, \ldots, n$$

This is a binary hypothesis test and the probability of error is given by

$$\underset{\xi}{Max} \quad \left\| \mathrm{diag}(\xi)\left(\boldsymbol{\varsigma}_1(\mathbf{x}, \hat{\mathbf{e}}_1) - \boldsymbol{\varsigma}_2(\mathbf{x}, \hat{\mathbf{e}}_2)\right)\right\|^2$$
$$where \quad \hat{\mathbf{e}}_1 = \arg\min \left\|\mathbf{diag}(\hat{\imath})\left(\mathbf{t} - \boldsymbol{\varsigma}_1(\mathbf{x}, \grave{\mathbf{e}}_1)\right)\right\|^2$$
$$\hat{\mathbf{e}}_2 = \arg\min \left\|\mathbf{diag}(\hat{\imath})\left(\mathbf{t} - \boldsymbol{\varsigma}_2(\mathbf{x}, \grave{\mathbf{e}}_2)\right)\right\|^2$$

Algorithm: T-Designs [1]

1. Two model case:

> 1. Given a design ξ_N, where N is the number of observations, find :
> $\hat{\theta}_{1N} = \arg\min_{\theta_1} \sum_{i=1}^{N} (t_i - \eta_1(x_i, \theta_1))^2$
> $\hat{\theta}_{2N} = \arg\min_{\theta_2} \sum_{i=1}^{N} (t_i - \eta_2(x_i, \theta_2))^2$
> 2. Add to the design a point x_{N+1} such that :
> $x_{N+1} = \arg\max_{x \in Z} \left(\eta_1(x_i, \hat{\theta}_{N1}) - \eta_2(x_i, \hat{\theta}_{2N})\right)^2$
> 3. The $(N+1)th$ observation is taken at x_{N+1}
> Update $\xi : \xi_{N+1} = (1-\alpha) * \xi_N + \alpha * \delta(x_{N+1})$
> 4. Go back to 1

2. Multiple model case:

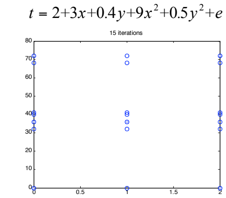
> 1. Given a design ξ_N, where N is the number of observations, for each model $\eta_k(x, \theta)$ find :
> $\hat{\theta}_{kN} = \arg\min_{\theta} \left(L_k = \sum_{i=1}^{N} (t_i - \eta_k(x_i, \theta))^2\right)$
> Rank the models by goodness of fit, for example
> $L_u \leq L_v \leq \ldots \leq L_w$
> 2. Add to the design a point x_{N+1} such that :
> $x_{N+1} = \arg\max_{x \in Z} \left(\eta_u(x_i, \hat{\theta}_{uN}) - \eta_v(x_i, \hat{\theta}_{vN})\right)^2$
> 3. The $(N+1)th$ observation is taken at x_{N+1}
> Update $\xi : \xi_{N+1} = \left(1 - \frac{1}{N+1}\right) * \xi_N + \frac{1}{N+1} * \delta(x_{N+1})$
> 4. Go back to 1

[1] A.C. Atkinson and V.V. Fedorov. Optimal design: Experiments for discriminating between several models. *Biometrika* 62, 289-303, 1975.

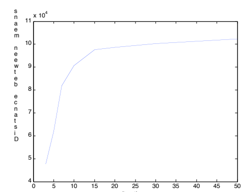
Evaluation:

$$M_1: t_i = \theta_{10} + \theta_{11}x + \theta_{12}y + e_i, \qquad i = 1, \ldots, n$$
$$M_2: t_i = \theta_{20} + \theta_{21}x + \theta_{22}y + \theta_{23}x^2 + \theta_{24}y^2 + e_i, \qquad i = 1, \ldots, n$$

$$t = 2 + 3x + 0.4y + 9x^2 + 0.5y^2 + e$$



Likelihood: $M_2 / M_1 = 2.18 * 10^5$



Future Work

Model Estimation:

- Optimal designs for *robust* model estimation.
- Optimal designs for multiple modality fields while incorporating the correlation between the modalities.
- Spatio-temporal fields.

Model Selection:

- Investigate the effect of the prior probabilities on the design.
- Extend the work presented to situations when the set of models considered does not include the correct model.

UCLA – UCR – Caltech – USC – UC Merced