Parameter Estimation in Autologistic Regression Models for Detection of Smoke in Satellite Images

Mark Wolters   Shanghai Center for Mathematical Sciences
Charmaine Dean   Western University

June 15, 2015
SSC 2015, Halifax

Outline

1. Introduction

Application; goals

2. Modelling Framework

CRF approach; autologistic regression; challenges

3. Approach to Estimation

Independence model + optimal smoothing

4. Some Results

Simulated data; preliminary findings

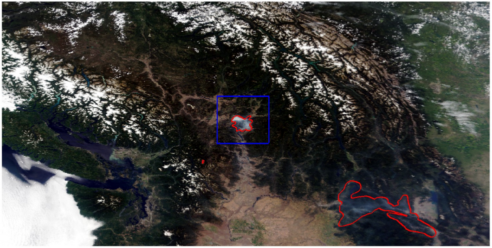
5. Discussion

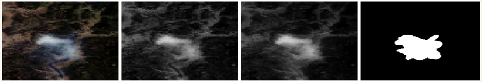
Possible improvements; estimation in a prediction setting

1. Introduction

Earth-orbiting satellites help study large-scale environmental phenomena.

Our interest: smoke from forest fires.





Original image Green Blue True class

Data: MODIS images

- 1 per day, 143 days
- 1.2 Mp each
- Centered at Kelowna, BC
- Hand-drawn smoke areas

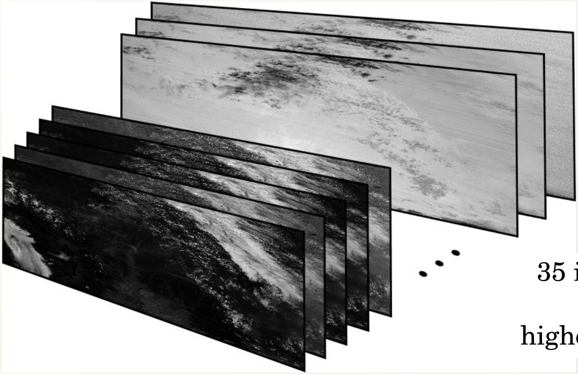
Goal: classify pixels into *smoke*/*nonsmoke*

Why?

- Health studies
- Model input or validation
- Monitoring & archiving

Data characteristics

- *Hyperspectral* images.
- Spectra at each pixel are covariates for predicting smoke.
- *High-dimensional* predictor space.
- Expect *spatial association*.



35 image planes
+
higher order terms

2. Modelling Framework

Joint PMF for an image: *autologistic regression* (ALR)

$$\Pr(\mathbf{Y} = \mathbf{y}) \propto \exp\left((\mathbf{X}\boldsymbol{\beta})^T \mathbf{y} + \frac{\lambda}{2} \mathbf{y}^T \mathbf{A} \mathbf{y} \right)$$

linear predictors are the *unary* coefficients

$\lambda =$ *pairwise* association parameter

$\mathbf{A} =$ adjacency matrix

$\mathbf{y}$	Class labels (n-vector), $y_i \in \{L, H\}$
$\mathbf{X}$	Model matrix (spectral data, $n \times p$)
π_i	P(pixel i is smoke | neighbour pixels)
$\mathcal{G} = (\mathcal{E}, \mathcal{V})$	Graph structure for dependence among pixels (regular 4-connected grid)

Autologistic regression

Notes:

- Intractable normalizing constant
- ALR is a *conditional random field* (CRF) model (Lafferty et al., 2001):
  - given $\mathbf{X}$, $\mathbf{Y}$ is a Markov random field.
- Conditional logit form:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = (H-L)\left(\mathbf{x}_i^T\boldsymbol{\beta} + \lambda\sum_{j\sim i} y_j\right)$$

- logistic regression $\iff \lambda = 0$
- Spatial effect is homogeneous, isotropic
- For $Y_i \in \{0,1\}$, the sum $\sum y_j$ increases log-odds unless all neighbours are zero.

  $\implies$ Estimates of $\boldsymbol{\beta}, \lambda$ are strongly coupled (Caragea and Kaiser, 2009; Hughes et al., 2011)

Extensions

- *Centered* ALR model (Caragea and Kaiser (2009)) aims to correct for the asymmetry of the pairwise term when $Y_i \in \{0, 1\}$:

$$\text{logit}(\pi_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \lambda \sum_{j \sim i} (y_j - \mu_j)$$

where $\mu_j = \text{E}[Y_j | \lambda = 0]$ is the independence expectation.

- **Claim:** just use $Y_i \in \{-1, +1\}$ to get the same effect.

- **Proposal:** let $\lambda = \lambda_{ij} = \lambda(\mathbf{x}_i, \mathbf{x}_j)$ for adaptive smoothing.

Then

$$\text{logit}(\pi_i) = 2 \left(\mathbf{x}_i^T \boldsymbol{\beta} + \sum_{j \sim i} \lambda(\mathbf{x}_i, \mathbf{x}_j) y_j \right)$$

3. Approach to Estimation

Existing possibilities

1. Ignore spatial association (logistic regression, large n, large p).
2. Pseudolikelihood (PL): $L(\boldsymbol{\beta}, \lambda) \approx \prod_{\text{img}} \prod_{i=1}^{n} \text{logit}(\pi_i)$
3. Monte Carlo ML
4. Bayesian approach

Hughes et al. use perfect sampling; recommend PL for large n. (applies to items 3 and 4)

Problems

- We have $\sim 10^8$ pixels
- We have thousands of predictors, need model selection
- We're still developing models—rapid evaluation of candidates is beneficial

Proposal: plug-in estimation

a) Use independence (logistic) to get $\hat{\boldsymbol{\beta}}$
   - Including model selection
   - Sample pixels to reduce n to manageable size

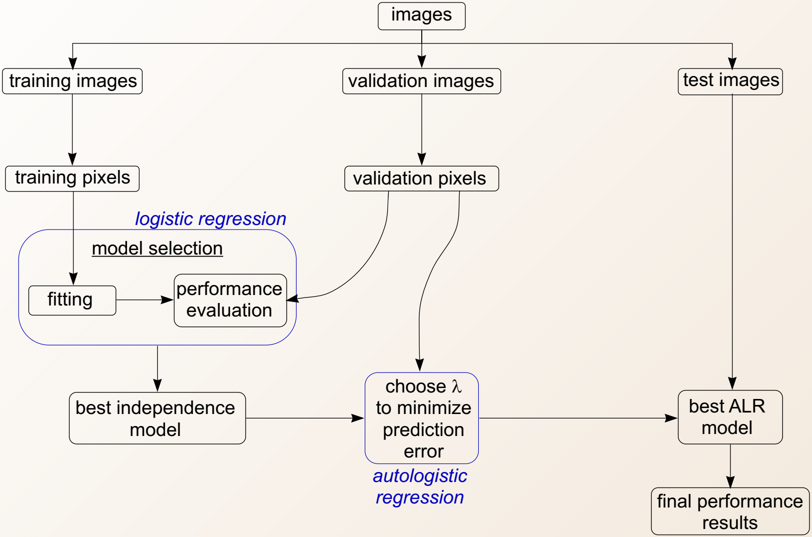
b) Choose $\hat{\lambda}$ to optimize predictive power

Rationale

Treat λ as a smoothing parameter.

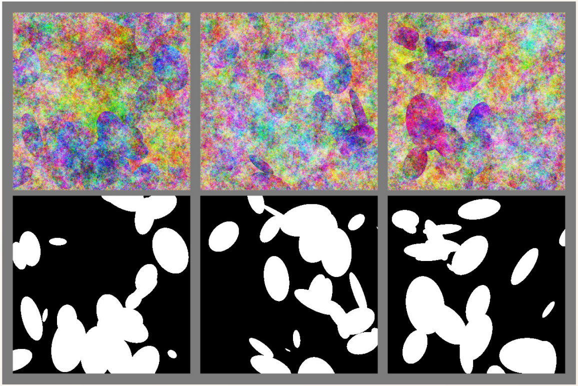
- Assuming independence, $\hat{\boldsymbol{\beta}}$ captures how information in $\mathbf{X}$ can be used to predict $\mathbf{Y}$.
- For fixed $\hat{\boldsymbol{\beta}}$, tuning λ will optimally reduce noise in the predicted probabilities.

Proposed procedure:



4. Results

Simulated Images



Predictors: R, G, B

Random ellipses = Class 1 (smoke)

Background = class 2 (nonsmoke)

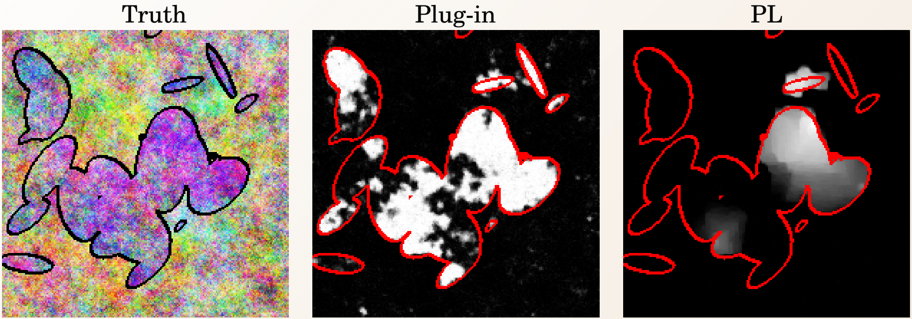
90 images at 3 sizes: 100, 200, 400 pixels square

A) plug-in vs. PL

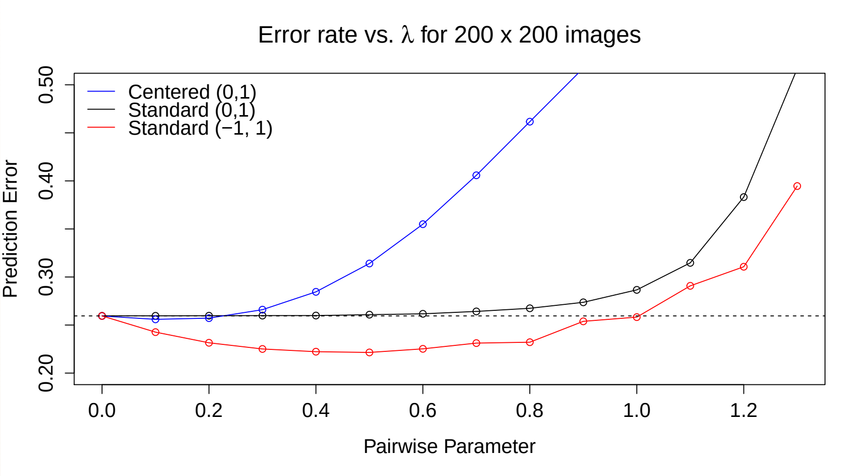
pixels	method	$\hat{R}$	$\hat{G}$	$\hat{B}$	$\hat{\lambda}$	error rate (%)	time (min)
100^2	plug-in	-2.20	-2.00	1.95	0.45	19.2	**1.2***
	PL	-2.04	-1.99	2.06	0.99	23.3	**0.9**
200^2	plug-in	-1.65	-1.36	1.71	0.5	20.8	**5.0***
	PL	-1.61	-1.30	1.70	1.19	48.2	**2.8**
400^2	plug-in	-2.00	-1.41	1.64	0.6	20.8	**21.5***
	PL	-2.08	-1.40	1.68	1.36	50.5	**16.1**

* time per candidate λ value

Example predicted probabilities



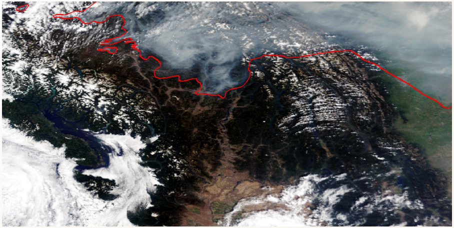
B) Effect of coding & centering



$$\text{logit}(\pi_i) = (H - L)\left(\mathbf{x}_i^T \boldsymbol{\beta} + \lambda \sum_{j \sim i} (y_j - \mu_j)\right)$$

C) Preliminary smoke results

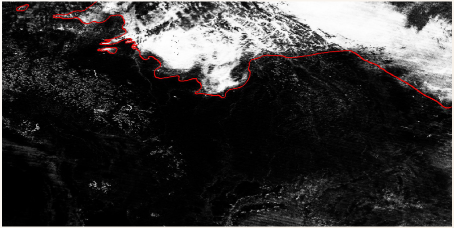
RGB image



ALR model with 50 predictors

1. Smoke-free areas: OK
2. Clouds vs. smoke: OK
3. Snow vs. smoke: OK
4. Spatial smoothing: OK
5. Smoke + Cloud: Problem

Logistic model

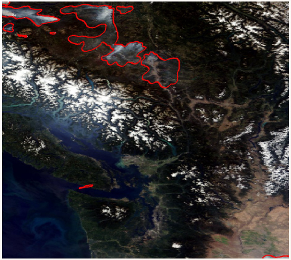


Autologistic model

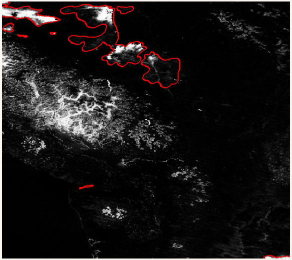
mwolters@fudan.edu.cn

Preliminary smoke results (continued)

RGB



Logistic



Autologistic

6. “Thin” smoke: Problem
7. Original masks (training data): Problem

5. Discussion

Advantages of working on a data-rich prediction problem

If you get low out-of-sample prediction error, the following are of little concern:

- The "truth" of your model
- The complexity or statistical efficiency of your model
- whether or not your parameters are statistically significant
- whether or not your parameter estimates are stable

Computational feasibility & run time become paramount.

How would things change if we're interested in interpretation?

- Trade predictive power for model simplicity.
- The "plug-in" estimation approach is no longer helpful.
- The issue of model centering and coding becomes critical.

Future plans: smoke

- Improve the base logistic regression model
- Address “true” label ambiguity

Future plans: models

- Computational improvements:
  - low-level code
  - parallelization
- Revisit adaptive smoothing
- A beta CRF for direct modelling of probabilities
- Multi-class (auto*binomial*) extension

References

Caragea, P. C. and Kaiser, M. S. (2009), "Autologistic models with interpretable parameters," *Journal of agricultural, biological, and environmental statistics*, 14, 281–300.

Hughes, J., Haran, M., and Caragea, P. C. (2011), "Autologistic models for binary data on a lattice," *Environmetrics*, 22, 857–871.

Lafferty, J. D., McCallum, A., and Pereira, F. C. N. (2001), "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," in *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 282–289.