CLOUD CLASSIFICATION OF SATELLITE RADIANCE DATA BY MULTICATEGORY SUPPORT VECTOR MACHINES

Yoonkyung Lee, Department of Statistics, The Ohio State University, Columbus
Grace Wahba, Department of Statistics, University of Wisconsin, Madison and
Steven A. Ackerman, Department of Atmospheric and Oceanic Sciences, University of Wisconsin, Madison

INTRODUCTION

- The MODIS (moderate resolution imaging spectroradiometer) measures radiances at 36 wavelengths with spatial resolution 250 m to 1 km.
- Earth Observing System (EOS) models require knowledge of whether a radiance profile is cloud free, or not.
- Support Vector Machines (SVM) can be useful as an efficient cloud detection and classification algorithm for EOS models.
- We have applied the Multicategory SVM to simulated MODIS type channels data to classify the radiance profiles as clear, liquid clouds, or ice clouds. The results are promising and better than channel by channel partitioning.

CLASSIFICATION PROBLEMS

- Training data set :
  Attribute $\mathbf{x}_i \in R^d$, and class label $y_i \in \{1, \cdots, k\}$, $i = 1, \cdots, n$.
  k is the number of classes.
- Task : Learn a classification rule matching $\mathbf{x}_i$ with y_i,
  the mapping $\phi(\mathbf{x}) : R^d \to \{1, \cdots, k\}$
- Example : Classify
  d-channel radiance profiles $\rightarrow$ {clear, liquid cloud, ice cloud}
  $\mathbf{x}$ {1, 2, 3}

SUPPORT VECTOR MACHINES

Overview

- Class label representation : $y_i \in \{-1, 1\}$.
- SVM looks for separating $f(\mathbf{x})$ with largest margin $yf(\mathbf{x})$ to match the sign of y.
- Classification rule : $\phi(\mathbf{x}) = sign(f(\mathbf{x}))$
- $f(\mathbf{x}) = h(\mathbf{x}) + b$ with $h \in \mathcal{H}_K$ a Reproducing Kernel Hilbert space (RKHS) with reproducing kernel $K(\cdot, \cdot)$.
  Linear SVM: $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$ where $\mathbf{w} \in R^d$ and $b \in R$.
  Nonlinear SVM: $f(\mathbf{x}) = \sum_{i=1}^n c_i K(\mathbf{x}_i, \mathbf{x}) + b$
- SVM scheme : to find f minimizing

$$\frac{1}{n}\sum_{i=1}^{n}(1 - y_i f(\mathbf{x}_i))_+ + \lambda \|h\|^2_{\mathcal{H}_K}$$

λ is a tuning parameter which balances the data fit and the complexity of $f(\mathbf{x})$.

Binary SVM and the Bayes Rule

- The minimizer of $E[(1 - Yf(X))_+]$ is $f(\mathbf{x}) = sign(p^+(\mathbf{x}) - 1/2)$, the optimal classification rule, where $p^+(\mathbf{x}) = P(Y = 1|X = \mathbf{x})$.
- SVMs implement the Bayes rule asymptotically when RKHS is rich enough to approximate $sign(p^+(\mathbf{x}) - 1/2)$.

MULTICATEGORY SUPPORT VECTOR MACHINES

Class codes

$\mathbf{y}_i$ is represented by a k-dimensional vector with 1 in the jth coordinate and $-\frac{1}{k-1}$ elsewhere if example i falls into class j.
For example, when $k = 3$,

$$\mathbf{y}_i = \begin{cases} (1, -\frac{1}{2}, -\frac{1}{2}) & \text{for class 1} \\ (-\frac{1}{2}, 1, -\frac{1}{2}) & \text{for class 2} \\ (-\frac{1}{2}, -\frac{1}{2}, 1) & \text{for class 3} \end{cases}$$

Separating functions

$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \cdots, f_k(\mathbf{x}))$ with sum-to-zero constraint, $\sum_{j=1}^k f_j(\mathbf{x}) = 0$ for any $\mathbf{x} \in R^d$, and $f_j(\mathbf{x}) = h_j(\mathbf{x}) + b_j$ with $h_j \in \mathcal{H}_K$.

Multicategory SVM formulation

- Find $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \cdots, f_k(\mathbf{x}))$, with sum-to-zero constraint, minimizing

$$\frac{1}{n}\sum_{i=1}^{n} L(\mathbf{y}_i) \cdot (\mathbf{f}(\mathbf{x}_i) - \mathbf{y}_i)_+ + \frac{\lambda}{2}\sum_{j=1}^{k} \|h_j\|^2_{\mathcal{H}_K}$$

  where $L(\mathbf{y}_i)$ = the jth row of the cost matrix C if $\mathbf{y}_i$ indicates class j. C is defined as k by k matrix with 0 on the diagonal, and 1 elsewhere. For example, when $k = 3$,

$$C = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

- The hinge loss $\mathcal{L}_{hinge}(\mathbf{y}, \mathbf{f}) = L(\mathbf{y}) \cdot (\mathbf{f} - \mathbf{y})_+$ forces f to match up the class code $\mathbf{y}$ in such a way that "too correct" f is encouraged.
- A classification rule induced from $\mathbf{f}(\mathbf{x})$: $\phi(\mathbf{x}) = arg\max_j f_j(\mathbf{x})$
- Binary SVM is a special case of Multicategory SVM when $k = 2$.

Multicategory SVM and the Bayes rule

- The Bayes rule: $\phi_B(\mathbf{x}) = arg\max_{j=1,\cdots,k} p_j(\mathbf{x})$,
  where $p_j(\mathbf{x}) = P(Y = j|X = \mathbf{x})$
- The minimizer of $E[L(Y) \cdot (\mathbf{f}(X) - Y)_+]$ under sum-to-zero constraint is $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \cdots, f_k(\mathbf{x}))$ with

$$f_j(\mathbf{x}) = \begin{cases} 1 & \text{if } j = arg\max_{\ell=1,\cdots,k} p_\ell(\mathbf{x}) \\ -\frac{1}{k-1} & \text{otherwise} \end{cases}$$

SIMULATED DATA

- Satellite observations at 12 wavelengths (.66, .86, .46, .55, 1.2, 1.6, 2.1, 6.6, 7.3, 8.6, 11, 12 microns or MODIS channels 1, 2, 3, 4, 5, 6, 7, 27, 28, 29, 31, 32) are simulated.
- A total of 744 radiance profiles over the ocean (81 clear scenes, 202 liquid clouds and 461 ice clouds). Each simulated radiance profile consists of 7 reflectances and 5 brightness temperatures.

Figure 1. **Box plots of 7 reflectances and 5 brightness temperatures**

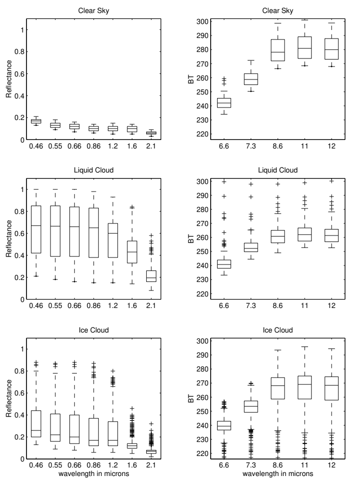


Figure 2. **10 randomly selected radiance profiles from each scene**

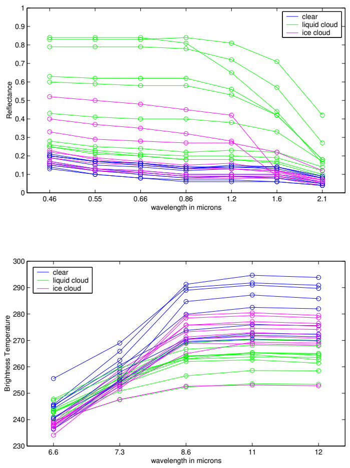
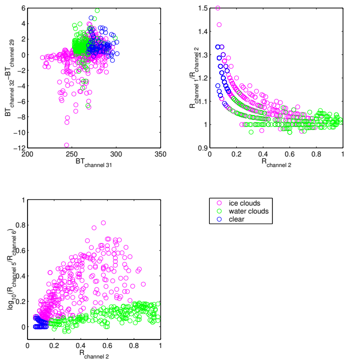


Figure 3. **Scatterplots of various features**



CLASSIFICATION RESULTS

Analysis set-up

- Split the data set into a training set and a test set. 370 profiles, almost half of the original data were selected randomly as the training set.
- To test how predictive the two features, $R_{channel_2}$ and $\log_{10}(R_{channel_5}/R_{channel_6})$ are, the MSVM with two features was applied to the training data first.
- Other features were added finally.
- Equal and unequal misclassification costs were considered.

Figure 4. **The classification boundaries determined by the MSVM using 370 training data**

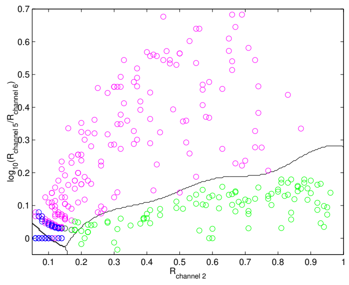
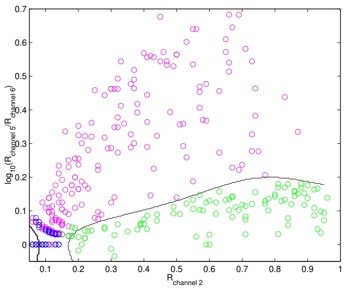


Figure 5. **The classification boundaries determined by the non-standard MSVM when the cost of misclassifying clouds as clear is 1.5 times higher than other types of misclassifications.**



Test error rates for the combinations of variables and classifiers

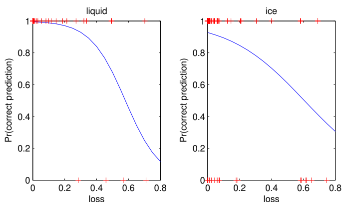
Number of variables	Variable descriptions	Test error rates (%) MSVM	TREE	1-NN
2	(i) R_2, $\log_{10}(R_5/R_6)$	11.50	14.97	16.58
5	(i)+R_1/R_2, BT_{31}, $BT_{32} - BT_{29}$	10.16	15.24	12.30
12	(ii) original 12 variables	12.03	16.84	20.86
12	log transformed (ii)	9.89	16.84	18.98

where TREE: a recursive partitioning method and
1-NN: the nearest neighbor

PREDICTION STRENGTH

- Attach a confidence measure to a class prediction.
- The hinge loss with the standard cost function $\mathcal{L}_{hinge}(\mathbf{y}, \mathbf{f})$ sensibly measures the proximity between an MSVM decision vector f and a coded class $\mathbf{y}$ in the classification context.
- Associate the hinge loss for the predicted class with the probability of a correct prediction given $\mathbf{f}(\mathbf{x}) = (f_1, \cdots, f_k)$ at $\mathbf{x}$, $P(Y = arg\max_j f_j|\mathbf{f})$.
- The strength of the MSVM prediction, $P(Y = arg\max_j f_j|\mathbf{f})$ can be inferred from a calibration set.

Figure 6. The estimated MSVM prediction accuracy as a function of the loss estimated via linear logistic regression, for the liquid and ice cloud predicted classes.



SUMMARY

- Multicategory SVM implements the optimal classification rule efficiently as the sample size gets large.
- The proposed method deals with the equal and the unequal misclassification costs in a unified way.
- The MSVM method clearly has benefits over the existing MODIS algorithms, which use thresholds on individual components of the attribute vectors.
- Although rudimentary, a procedure of assessing the strength of the prediction is proposed.
- It has a great potential for applications in atmospheric sciences.

REFERENCES

- Ackerman, S. A., Strabala, K. I., Menzel, W.P., Frey, R. A., Moeller, C.C., and Gumley, L.E. (1998) *Discriminating Clear-sky from Clouds with MODIS*, Journal of Geophysical Research, **103**, 32, 141–157.
- Lee, Y., Lin, Y., and Wahba, G. (2002) *Multicategory Support Vector Machines, Theory, and Application to the Classification of Microarray Data and Satellite Radiance Data*, Technical Report 1064. Department of Statistics, University of Wisconsin-Madison.
- Lee, Y., Wahba, G., and Ackerman, S. A. (2003) *Cloud classification of Satellite Radiance Data by Multicategory Support Vector Machines*, Technical Report 1075. Department of Statistics, University of Wisconsin-Madison.
  Papers are available from
  http://www.stat.wisc.edu/~wahba/trindex.html, or
  http://www.stat.ohio-state.edu/~yklee/paper.html.
- Strabala, K. I., Ackerman, S. A., and Menzel, W.P. (1994) *Cloud properties inferred from 8-12-μm data*, Journal of Applied Meteorology, **33**, 212–229.

ACKNOWLEDGMENTS

This research was partly supported by NASA Grant NAG5 10273 and NSF Grant DMS0072292.