Detection of Sangeang Api Volcanic Ash Cloud Based on VBICA-SVM method

Jiangshan DONG, Chengfan LI*, Jingyuan YIN, Junjuan ZHAO, Dan XUE School of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China *corresponding author email: david-0904@163.com

Keywords: Remote sensing image; Independent component analysis (ICA); Support vector machine (SVM); Volcanic ash cloud

Abstract. Volcanic ash cloud can threaten the safety of aviation transportation. The current detection of the volcanic ash cloud mainly depends on the remote sensing images. Independent component analysis (ICA) and support vector machine (SVM) have very strong complementary to each other in the volcanic ash cloud detection. Taking the moderate resolution imaging spectroradiometer (MODIS) images of Indonesia Sangeang Api volcanic ash clouds as data source on May 30, 2014, this paper propose the variable Bayesian ICA and SVM (VBICA-SVM) method for detection of the volcanic ash cloud, which is based on the analysis of the physical properties of volcanic ash cloud and the absorption spectral characteristics in the scope of thermal infrared bands. The result shows that the VBICA-SVM method can effectively detect the spatial distribution of the volcanic ash clouds, and has great potential applications in the ash cloud monitoring and disaster mitigation.

Introduction

Remote sensing has the advantages of wide coverage, large amount of information and short revisit cycle, and widely used in the resource investigation, environmental monitoring, etc. [1, 2] Compared with the traditional digital image, the remote sensing images contain more diverse terrain types, therefore, the objects distribution in remote sensing image is more complex. The ash cloud monitoring using remote sensing technique is a huge system project, and the key is how to accurately identify the spatial distribution of the ash cloud [3, 4]. Although the researchers at home and abroad used all kinds of remote sensing data and methods for identification study on some large-scale volcanic ash cloud [5], most of these methods are based on the mathematical model, and can't completely meet the requirements of actual ash cloud detection.

As a new signal processing technology, ICA has close association with the problem of blind source separation, and it has the ability of separating source signals from the mixed ones without the need of prior knowledge [6, 7]. The SVM can realize the non-linear classification of objects in remote sensing image. Both have a great advantage in remote sensing images processing. At present, the application field of the ICA and SVM has expanded from solution of mixed blind signals to signal detection, image processing and pattern recognition, etc. But the ICA model has problems containing the independence and invariance of components, no noise assumption as well as the uncertainty of the final solution [8], and SVM has problem such as small sample properties. Given this, this paper proposes VBICA-SVM method used for detection of volcanic ash cloud from remote sensing images. In the proposed method, the Bayesian network is introduced into the ICA model, we used Bayesian inference to complete the study of unknown hidden variables (independent components), and we optimized it by combination with the variational approximation algorithm to make the isolated independent component as consistent as the real situation of the surface. And then the obtained independent component of VBICA make as the inputted feature vector of SVM.

The first section of our paper is the introduction. The second section is the theoretical basis of ICA and SVM method. The third section is the detection of Indonesia Sangeang Api volcanic ash cloud based on VBICA-SVM method. The fourth section is precision evaluation of Indonesia

Sangeang Api volcanic ash cloud, and the last section is our conclusion.

Theoretical basis

Add noise information in the standard ICA model, making it the linear mixed ICA model with noise [9, 10], and the formula is shown below:

 $x(t) = As(t) + \varepsilon(t)$

(1)

Where the x(t) refers to a mixed signal with M dimension, s(t) refers to a source signal (hidden variable) with L dimension, mixed matrix A is $M \times L$, $\varepsilon(t)$ refers to Gaussian noise, which usually is a diagonal matrix with the inverse variance for Λ and the mean for zero.

In linear mixed ICA model containing noises, the probability calculation formula for mixed signal x(t) is:

$$p(x \mid s, A, \Lambda) = \left| \det(\frac{1}{2\pi}\Lambda) \right|^{\frac{1}{2}} \exp[-E_D]$$
(2)

Where $E_D = \frac{1}{2}(x - As)^T \Lambda(x - As)$, det(·) refers to the determinant value. In the model we assume

the source signals are independent of each other, so the probability distribution formula for s can be expressed as $p(s) = \prod_{i=1}^{L} p(s_i)$.

In Bayesian network model, all the parameters and the hidden variables of the model are represented by $W, W = \{A, A, S, q, \theta\}$. The maximum negative free energy $F[x|\theta]$ is the objective function of the model, the formula of which is shown by the following:

$$F[x \mid \theta] = \left\langle \log p(x, W \mid M) \right\rangle_{p'(W \mid \theta)} + H[p'(W)]$$
(3)

In variational Bayesian ICA model, according to the conditional independence principle of the Bayesian network, we use the approximate form of real posteriori probability distribution $p'(W|\theta)$ to calculate the partial derivatives of various parameters, and the maximum value, respectively, so:

$$p'(W_i \mid \theta) = \frac{1}{Z_i} \exp[\langle \log(x, W \mid M) \rangle_{\prod_{i \neq j} p'(W_{j \neq i} \mid \theta_{j \neq i})}]$$
(4)

Where Z_i refers to specification factors, whose value varies with the variable.

In variational Bayesian ICA model, through the conditional independence assumption of Bayesian inference and source signal generation model, we can choose the approximate form of real posteriori probability distribution $p'(W|\theta)$:

$$p'(W \mid \theta) = p'(\Lambda)p'(A)p'(s \mid q)p'(q)p'(\theta)$$
(5)

Where $p'(\theta) = p'(\pi)p'(\mu)p'(\beta)$, p'(s | q) can be obtained in GOM model.

According to the source signal generation model of variational Bayesian ICA, the following formula can be launched:

 $p(x|W) = p(x|A,\Lambda,s)p(s|q,\mu,\beta)p(q|\pi)p(\theta)p(\Lambda)p(\Lambda)$ (6)

Substitute the approximate form of real posteriori probability distribution $p'(W|\theta)$ into maximum negative free energy *F*. Combining with the definition of prior distribution of the parameters in the source signal model, the best approximation true posterior probability distribution parameters of the source signal *s* can be deduced:

$$p(S \mid q) = \prod_{i=1}^{L} \prod_{t=1}^{T} N(S_i^t; \hat{\mu}_{i,q_i}^{(t)}; \hat{\beta}_{i,q_i}^{(t)})$$
(7)

Similarly, we can also continuously deduce the real posterior probability distribution of other parameters and hidden variables in the variational Bayesian ICA model, and obtain the approximation value through iteratively approaching true value until convergence.

Suppose there is a sample set $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, x_i \in R_d$ represent the input mode, $y \in \{\pm 1\}$ represent targets output, $i \in (1, 2, \dots, N)$. The optimal hyperplanes for classification is $w_{Txi} + b = 0$ and the weight *w* and bias *b* should satisfy the following conditions:

$$y_i(w^T x_i + b) \ge 1 - \xi_i \tag{8}$$

Where ξ_i is a slack variable indicating the schema from the linearity, $i = 1, 2, \dots N$. SVM aims to find a hyperplane to make the average classification error of training data reach minimum and thereby:

$$\Phi(w,\xi) = \frac{1}{2}w^T w + C \sum_{i=1}^{N} \xi_i$$
(9)

Where *C* is the penalty coefficient indicating the punishment degree of error-classifying samples by SVM, $C \ge 0$. On the basis of Lagrangian, the solution of optimal hyperplanes for classification can be converted into the following formula:

$$Q(a) = \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j y_j y_j K(x_i, x_j)$$
(10)

Where $\{a_i\}_{i=1}^N$ is the Lagrangian multiplier, the vast majority of which abide by $a_i = 0$ and the constraint condition is $a_i > 0$. The corresponding samples will be support vectors only when $a_i \neq 0$, $K(x_i, x_j)$ is the SVM's kernel function that satisfies Mercer theorem.

Detection of Sangeang Api volcanic ash cloud

General situation

Sangeang Api, also known as Abigail volcano, is located in Indonesia, and it's also a volcanic island. It is a part of the Lesser Sunda Islands, and it is governed by the West Nusa Tenggara. It is a stratovolcano composed by two volcanic cones, the altitudes of which are 1795 m and 1949 m respectively, and it is one of the most active volcanos in Lesser Sunda Islands. The ash erupted in May 30, 2014 reached up to 15-20 km, flew over the ocean, arrived in the Kimberley region in the northwest of Australia and affected many flights in the local area. Figure 1(a) shows the partial enlarged image of MODIS image of Indonesia Sangeang Api volcanic ash cloud on May 30, 2014.

This paper uses the Terra/MODIS satellite remote sensing data of the eruption in May 30, 2014 to detect and study volcanic ash clouds. Figure 1(b) is MODIS false color composite image of bands 2, 1, 3 on May 30, 2014, and the center wavelengths of three corresponding bands are $0.859\mu m$, $0.645\mu m$ and $0.469\mu m$ respectively. From Figure 1(b), the volcanic ash cloud presents the hoary in the MODIS image, and the difference from the sea water and islands is obvious. However, it confused with the meteorological cloud seriously.



Fig.1. The partial enlargeed MODIS image of Indonesia Sangeang Api volcanic ash cloud (a) and MODIS false color composite image (b)

The detection of Sangeang volcano ash cloud

We can learn from the basic principle of PCA that after PCA processing the first principal component images (PCI1) has the largest amount of information, concentrated with around 80% of the whole ash cloud information, and it is followed by the second principal component images (PCI2), then the principal component images (PCI) show a decreasing amount of information in

turn. In our experiment, we handle the outer band data (except band 26) on the ash cloud of Sangeang volcano on May 30, 2014 with the help of the remote sensing image processing software ENVI 4.6 respectively, and the result is shown in Figure 2.



Fig.2. The MODIS principal component images of Sangeang volcano ash cloud, (a) – (d) refer to PCI1 – PCI4 respectively, PCI5 - PCI16 are omitted for there are basically all noise.

Seen from Figure 2, the color of the volcanic ash cloud and meteorological cloud is darker and film in PCI1 image, while the color of the sea and land is mostly gray, the boundary contrast of volcanic ash cloud and meteorological cloud with the land, sea and other types of surface feathers is very obvious as a whole. Then we see PCI2, after distinguish the volcanic ash cloud and meteorological cloud information from PCI image, PCI2 image highlights part of the pieces of information on the southeast volcanic ash cloud to a certain extent. In PCI3 image, although the ash cloud information is seperated from the meteorological cloud, sea and land, the volcanic ash information become vague and confused with some coasts of bare islands. In short, after PCA processing for the infrared MODIS data, the information of the volcanic ash cloud, meteorological cloud, land, sea and other different types of surface feathers has been gathered, which highlights the ash cloud information and raises the separability of the cloud. The PCA processing is helpful for the subsequent ash cloud recognition research based on PCI image to a certain extent.

Figure 3 is the independent component image (ICI) for Sangeang volcano ash cloud after VBICA processing.



Fig.3. MODIS independent component image for Sangeang volcano ash cloud obtained by the variational Bayesian ICA method, (a) - (p) refer to ICI1 - ICI16

Seen from Figure 3, in the ICI image which is processed with VBICA method, the difference between information in different terrain type increases obviously, i.e. the separability between different terrain types is increased, which is propitious to help improve the extraction accuracy of volcanic ash cloud thematic information. According to visual interpretation, ICI3 image has the largest amount of volcanic ash cloud information, and the contrast of ash cloud information with meteorological cloud, sea and land is very clear. Other ICI images either have the disadvantage of poor contrast between different components, or have the deficiency that the misjudgment phenomenon for the volcanic ash cloud information is serious.

In addition, in order to understand the quality of ICI image after being processed by the variational Bayesian ICA method more detailedly, we calculate the color histogram for every ICI image. Results show that in ICI images after VBICA processing, the difference between the ash cloud information and the sea, land and other background features increase, and its spectrum

distribution is more even, to a certain extent, it increases the separability of volcanic ash cloud thematic information in MODIS remote sensing images.

We use ICI images on Sangeang volcano ash cloud after being processed by VBICA method as the input feature vectors of the SVM, participating in the SVM classifier training and learning, respectively. We choose RBF kernel function as the kernel function of SVM classifier, the penalty coefficient *C* is 800, and the interval coefficient γ is 0.035. Figure 4 is the Sangeang volcanic ash cloud information on May 30, 2014 detected by VBICA - SVM.



Fig.4. the Sangeang volcanic ash cloud information on May 30, 2014 detected by VBICA – SVM method, the white part is the volcanic ash cloud.

Seen from Figure 4, VBICA-SVM method accurately identified the Sangeang volcanic ash cloud information on May 30, 2014, the broken patches of the identified volcanic ash cloud information are very few, and the visual effect is good.

The precision evaluation

To evaluate the recognition effect of the ash cloud information obtained by VBICA-SVM method, in this experiment we randomly generate 500 calibration data points, then we the confusion matrix method is adopted to help conduct the accuracy evaluation for volcanic ash cloud information, the results are shown in Table 1 and Table 2.

		volcanic ash clo	oud non-volcanic ash clo	ud total	
	volcanic ash cloud	113	31	144	
	non-volcanic ash clou	ıd 27	329	356	
	total	140	360	500	
	Table.2. The ider	ntification accuracy	for the ash cloud inform	nation	
	reference pixels number	classification pixels number	right classification pixels number	producer precision	user precision
volcanic ash cloud	140	144	113	80.72%	78.47%
non-volcanic ash cloud	360	356	329	91.39%	92.42%
total	500	500	442	_	
	total precision $= 88.4$	-%	Kappa coeffic	ient = 0.801	1

Table.1. The confusion matrix for the ash cloud information

As we can see from Table 1 and Table 2, the total accuracy of identified ash cloud information obtained by VBICA-SVM method reaches 88.4%, and the Kappa coefficient is 0.8011, thus good performance has been achieved.

Conclusions

This paper applied VBICA-SVM method on the identification research of the Sangeang volcanic

ash cloud information on May 30, 2014 in MODIS images, in terms of both recognition accuracy and image quality, the identification effect of ash cloud information is good enough. In addition, the algorithm can overcome the problem of being not able to distinguish the ash cloud information with meteorological cloud, which appears in the above two methods, thus it can be used as a useful supplement for traditional ash cloud recognition methods. However, the method also has the problem of time-consuming calculation and computing complexity. In the subsequent research, more attention should be paid on tring to reduce the parameters participated in variational Bayesian ICA calculation as much as possible, and on the problem of ranking the separated independent components according to certain rules.

Acknowledgement

This work was co-supported by the Projects of National Science Foundation of China (41404024), Shanghai Science and Technology Development Foundation (14231202600), Laboratory Technician Team Building Program in Shanghai Universities (B.60-E108-14-101) and Young Teachers Training and Supporting Plan in Shanghai Universities (2014-2016).

References

[1] Ellrod G.P. & Schreiner A.J., Volcanic ash detection and cloud top height estimates from the GOES-12 imager: Coping without a 12 μ m infrared band[J]. Geophysical Research Letter, 2004, 31(15), 1-4.

[2] Gangale G., Prata A.J. & Clarisse L., The infrared spectral signature of volcanic ash determined from high-spectral resolution satellite measurements[J]. Remote Sensing Environment, 2010, 114(2), 414-425.

[3] Papp K.R., Dean K.G. & Dehn J., Predicting regions susceptible to high concentrations of airborne volcanic ash in the North Pacific region[J]. Journal of Volcano Geothematic Research, 2005, 148(3-4), 295-314.

[4] Prata A.J., Satellite detection of hazardous volcanic clouds and the risk to global air traffic[J]. Natural Hazard, 2009, 51(2), 303-324.

[5] Prata A.J., Observations of volcanic ash clouds in the 10-12µm window using AVHRR/2 data[J]. International Journal of Remote Sensing, 1989, 10(4-5), 751-761.

[6] Gary P.E., Impact on volcanic ash detection caused by the loss of the 12.0 μm Split Window band on GOES imagers[J]. Journal of Volcano Geothematical Research, 2004, 135(1-2), 91-103.

[7] Ding F. & Xu H.Q., Sensitivity analysis of mono-window and single-channel algorithms to the possible errors in parameter estimation[J]. Science of Survey Mapping, 2007, 32(1), 87-95.

[8] Qin Z.H., Karnieli A. & Berliner P., A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel-Egypt Border Region[J]. International Journal of Remote Sensing, 2001, 22(18), 3719-3746.

[9] Price J.C., Land surface temperature measurements from the split window channels of the NOAA 7 advanced very high resolution radiometer[J]. Journal of Geothematical Research, 1984, 89(D5), 231-237.

[10] Becker F. & Li Z., Towards a local split window method over land surface[J]. International Journal of Remote Sensing, 1990, 11(3), 369-393.