The application of remote sensing image sea ice monitoring method in Bohai Bay based on C4.5 decision tree algorithm

Wei YE, Wei SONG*

Tianjin Research Institute for Water Transport Engineering, M.O.T., Tianjin 300456, China

*Corresponding author: S56wei@gmail.com

Abstract. In The Paper, the remote sensing monitoring of sea ice problem was turned into a classification problem in data mining. Based on the statistic of the related band data of HJ1B remote sensing images, the main bands of HJ1B images related with the reflectance of seawater and sea ice were found. On the basis, the decision tree rules for sea ice monitoring were constructed by the related bands found above, and then the rules were applied to Liaodong Bay area seriously covered by sea ice for sea ice monitoring. The result proved that the method is effective.

1. Introduction

As early as 1992, Key and Haefliger pointed out that land satellites could be used for positioning studies of sea ice activity and trend. There have been many research methods related to sea ice monitoring by satellite remote sensing, is mainly SAR (Synthetic Aperture Radar) and MODIS (Moderate-resolution Imaging Spectroradiometer) as the data source for sea ice monitoring. Han Suqin et al. [1] Extracted the distribution of sea ice and the characteristic values of the outer edge of sea ice using the characteristics of sea ice reflected in visible, near infrared and far infrared channels, and the results show that a large extent of sea ice can be detected by using MODIS remote sensing images. Wu Kuiqiao [2], Wu Longtao [3] et al. carried on the sea ice parameter inversion mainly using the MODIS remote sensing data, and provide remote sensing images of sea ice and numerical products such as sea ice concentration. Franz, J.Meyer[4] et al. extracted the range of the offshore sea ice by processing the interferometric phase pattern and the interferometric coherent images based on the L band SAR data. Sungwook Hong[5] et al. retrieved the roughness coefficients of small scale sea ice and the refractive index of sea ice by passive microwave, and extracted sea ice by the refractive index.

The essence of sea ice monitoring is to classify sea ice and sea water as two categories. Data mining techniques for classification problems have inherent advantages and are capable of discovering potentially useful and unknown rules and knowledge. So far, many researchers have adopted data mining techniques to solve classification problems. Using the ASTER remote sensing data as the data source, Li Mingshi [6] trained the datasets for 8 main terrain types, and extracted the spatial distribution information of these terrain types respectively using the maximum likelihood method, BP neural network method and decision tree classification algorithm. By comparison, the result showed that the decision tree classification algorithm has the best classification performance. Wang Changying [7] conducted in-depth discussion on the remote sensing image classification method of coastal zone based on Data Mining.

The sea ice of Liaodong Bay Area being heaviest icing area in Bohai Bay was monitored and the range of sea ice was extracted by decision tree classification method, using HJ1B remote sensing images as data source. At Last the precision of the extraction was verified.

2. Research area and data

2.1. Research area
Bohai sea is a part of the Western Pacific Ocean and also an inland sea of china. Bohai sea is composed of five parts, including the northern Liaodong Bay, western Bohai bay, south of Laizhou bay, central shallow sea basin and Bohai strait. During the year, sea ice appeared only in winter in Bohai sea and the north of the Yellow Sea. The glacial period of Bohai sea and the northern of the Yellow Sea is about 3~4 months. The period of the Liaodong bay, the north of the Yellow Sea, Bohai Bay and Liaodong Bay shortens in turn.

2.2. Research data
The HJ1B remote sensing images of Bohai Sea in February 8, 2010 and February 4, 2012 are the research data of this paper. The decision tree rules were constructed by experiments on the images in February 8, 2010, and then the validity of the C4.5 decision tree model is tested on the images in February 4, 2012.

3. The remote sensing image sea ice monitoring method based on the decision tree algorithm

3.1. Data pre-process
The CCD images from HJ1 satellite were preprocessed firstly, including radiometric calibration, geometric correction, registration of CCD1 and CCD2 data, data mosaic, data cutting and so on.

3.2. Construction of the sample table
5 CCD images with better imaging quality of Liaodong Bay in February 8, 2010 were as the training sample data set. In order to extract the original features of HJ1B image data, 35560 ice covered pixel samples and 21116 ice free pixel samples were selected for sample table after geometric correction. The data format of the sample table was shown in table 1.

Table 1. The sample table.

ID	X	Y	B1	B2	B3	B4
1	160	40	0.0389	0.0654	0.145	0.0068
2	160	41	0.0392	0.0652	0.143	0.0068
...	…	…	…	…	…	…

3.3. Statistical analysis of sample data and collation of training samples
Statistical analysis was carried out for the data selected above. Firstly, the mean values of the 4 bands of sea ice and sea water are calculated and shown in figure 1.
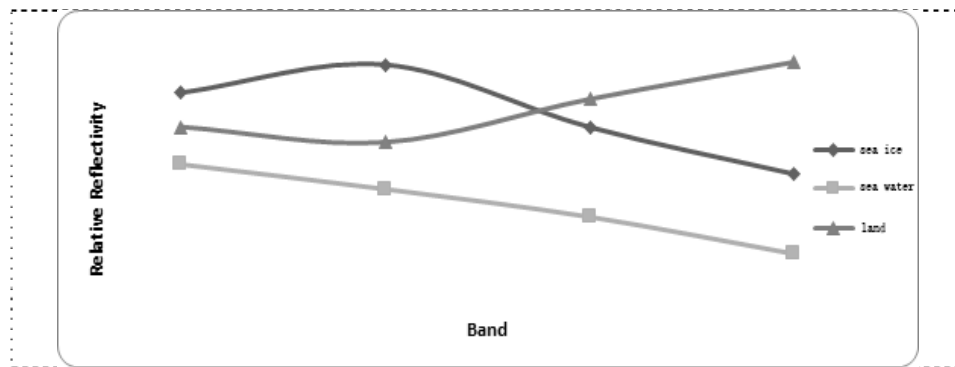
Figure 1. The relative reflectivity of each band of sea ice, sea water and land.

As shown in Figure 1, the reflectivity of sea ice and sea water is very different from the average in the B1 and B2 band. Figure 2 shows the statistical analysis of the reflectivity of sea ice and sea water in the B1 and B2 band respectively. The results show that the reflectivity of sea ice in the B1 band is mostly concentrated at 0.05~0.15, while the reflectivity of sea water in the B1 band is mainly concentrated on 0.02~0.05. Based on the result, the B1 bands are chosen as characteristic attributes for separating sea ice and sea water. Similarly, from Figure 1, the reflectivity of sea ice and land was found to differ considerably in the mean values of the B2 and B4 band.
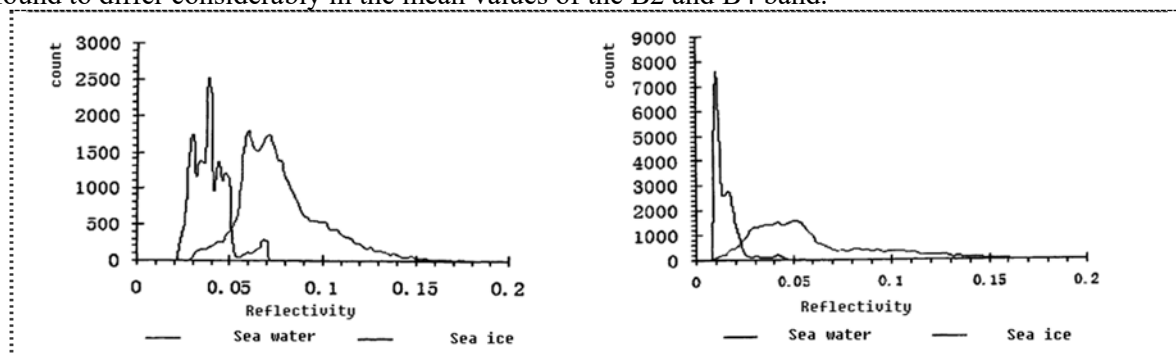

Figure 2. The frequency of the reflectivity of sea ice and sea water in B1 and B2 wavelengths.

In addition, NDSI [8] (Normalized Difference Snow Index) is used to extract the information of snow, mainly due to the characteristics of snow with high reflectivity in the visible band and with low reflectivity in short wave infrared. And sea ice has similar characteristics. The formula for NDSI is:

NDSI=(B3-B4)/(B3+B4)

There are three relationships between sea ice and background data: sea ice and sea water; sea ice and land; sea ice, sea water and land. To validate the C4.5 decision tree algorithm, the paper divided the experiment into three sets of training sample data.

The sample set selected in the 3.2 section was used as the training sample. the NDSI was calculated according to the formula. Then B1 and B2 band were selected as a set. B2 and B4 band were selected as a set. B1, B2 and B4 were used as the characteristic attributes together, and then a Category attribute is added. The complete data structure of the training sample set is shown in table 2.

Table 2. The training sample data structure.

Experiment sets	Characteristic attributes				
Sea ice and sea water	B1	B2		NDSI	CLASS
Sea ice and land		B2	B4	NDSI	CLASS
Sea ice, sea water and land	B1	B2	B4	NDSI	CLASS

3.4. Construction of decision tree model
C4.5 [9] algorithm is a classic decision tree algorithm is developed based on the ID3 algorithm. Due to the use of cross validation (Cross Validation) [10], C4.5 algorithm only requires a training set, does not need special test set. The tree display of the C4.5 decision tree classification model constructed finally in the paper was shown in figure 3.
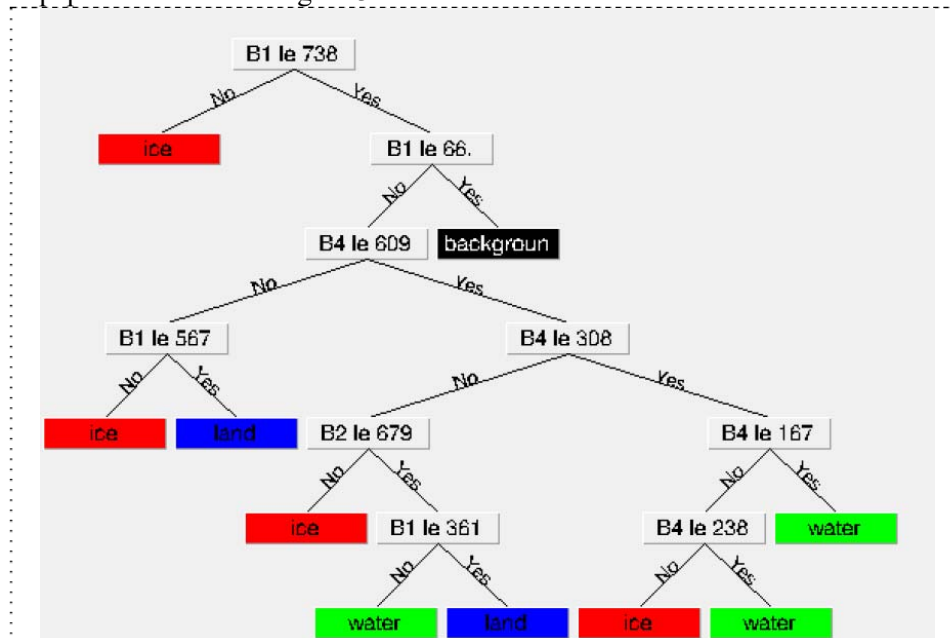

Figure 3. Sea ice, sea water and land decision tree classification model.

3.5. Validation of decision tree model
Based on the 5 Scenes HJ-1B CCD data of Bohai sea in February 4, 2012, which is of better imaging quality, the decision tree classification model constructed above was applied to validate the effectiveness of sea ice monitoring. There was the HJ-1B image after pre-process on the left of figure 4, and the sea ice monitoring result map generated by the model trained in the previous section on the right. The sea area covered by ice was red, while the land area was black, and the sea area not covered by ice was dark green.

Figure 4. HJ image(left) and sea ice monitoring map(right) of Liaodong Bay.

The accuracy of the above sea ice monitoring results was evaluated by using the error matrix, and the result was shown in table 3. Using C4.5 decision tree method, the extracted sea ice area of Liaodong Bay was 15243.831 square kilometer. While the satellite remote sensing monitoring result released by the Meteorological Bureau of Liaoning province in 2012 was 16391.216 square kilometer. The accuracy of the C4.5 decision tree model in the paper was up to 93%.

Table 3. C4.5 sea ice monitoring classification accuracy evaluation.

	Fixed ice	Floating ice	Land	Sea water 1	Sea water 2	Sea water 3	User accuracy%
Fixed ice	3858		6	492	102	516	76.56
Floating ice	78	4014					98.19
Land	24		630	7			96.45
Sea water 1	80		8	4122			97.81
Sea water 2	156				2478		94.18
Sea water 3	48					1044	96.60
Producer's accuracy%	90.60	100.00	99.23	89.22	97.05	71.92	

Note: the total accuracy of classification is 90.9701%, and the coefficient of Kappa is 0.8903.

4. Conclusion and suggestion

4.1. Conclusion

In this paper, the problem of sea ice monitoring in remote sensing images is transformed into data mining classification problem. Through the statistics, analysis and necessary operations of remote sensing image data, a decision tree for sea ice monitoring is constructed by using decision tree algorithm, and the rules of sea ice monitoring in HJ images are obtained. Experiments were carried out using the February 8, 2010 Bohai sea remote sensing images, the results of the C4.5 decision tree algorithm are compared with the results of the ISODATA method, and the relevant rules were applied to the February 4, 2012 sea ice monitoring remote sensing image of Bohai sea in ice serious Liaodong Bay Area. The area of the sea ice in Liaodong Bay area by C4.5 decision tree extraction is 15243.831 square kilometres, compared with 16391.216 square kilometres of satellite remote sensing monitoring results released by Liaoning Meteorological Administration in 2012, the accuracy rate reached 93%. The validity of the method is proved.

4.2. Suggestion

(1) Sea ice information extraction is performed using techniques other than decision tree classification techniques, constantly improve the sea ice information extraction technology to achieve higher accuracy.

(2) The decision tree is used to classify sea ice information and apply it to other remote sensing images, such as TM, MODIS and so on.

References

[1]   HAN Suqin, LI Zhen FA and SUN Zhigui 2005 Observational Study Of MODIS Data On Sea Ice Of China's Bohai Sea *J. Scientia Meteorologica Sinica* **25(6)** pp 624-628.

[2]   WU Qiaokui, XU Ying and HAO Yimeng 2005 Application In Sea Ice Remote Sensing Of MODIS Data *J. Marine Forecasts* **22(z1)** pp 44-49.

[3]   WU Longtao, SUN Lantao and SUN Lantao 2006 Retrieval of Sea Ice in the Bohai Sea from MODIS Data *J. Periodical Of Ocean University Of China* **36 (2)** pp 173-179.

[4]   FJ Meyer, AR Mahoney and H Eicken 2011 Mapping Arctic Landfast ice Extent Using L-Band Synthetic Aperture Radar Interferometry *J. Remote Sensing of Enviroment* **115(12)** pp 3029-3043.

[5]   Sungwook Hong 2010 Detection of Small-Scale Roughness and Refractive Index of Sea Ice in Passive Satellite Microwave Remote Sensing *J. Remote Sensing of Environment* **114(5)** pp 1136-1140.

[6]   LI Mingshi, PENG Shi-kui and ZHOU Lin 2006 A Study Of Automated Construction And Classification Of Decision Tree Classifiers Based On Aster Remotely Sensed Datasets *J. Remote Sensing For Land & Resources* **18(3)** pp 33-36.

[7]   C Wang, J Zhang and Y Ma 2008 Coastal Land Covers Classification of High-Resolution Images Based on Dempster-Shafer Evidence Theory *A. International Conference On Computer Science & software engineering* **1** pp 1061-1064.

[8]   LI Jun, LI Yan-hui and CHEN Shuang-ping 2009 Fast image denoising with CUDA *J. Computer Engineering And Applications* **45(11)** pp 183-185.

[9]   DAI Nan 2003 Research on classification methods based on decision tree algorithm *D. Nanjing, Nanjing Normal University*.

[10]   XU Jianfeng, SHAO Fengjing and YU Zhongqing 2002 A Rule Extraction Algorithm Base on the CART Classifier Pro *C. the 8th Joint International Computer Conference* **1** pp 545-547.