International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.9, (2016), pp.155-166
http://dx.doi.org/10.14257/ijsip.2016.9.9.15

Ensemble Estimation of Aerosol Optical Depth by Feature Selections from Remote Sensing Data

Bo Han

International School of Software, Wuhan University, Wuhan, 430079, China
bhan@whu.edu.cn

Abstract

Aerosol optical depth (AOD) is an important quantity parameter to study the Earth's radiation balance, climate change and environment protection. For estimating AOD by a data mining method, the synchronized records by combing satellite observed information from MOderate Resolution Imaging Spectroradiometer (MODIS) equipment with the ground-based accurate measurements of AOD from Aerosol Robotic NETwork (AERONET) work as driving attributes and prediction targets, respectively. However, compared with the number of high-dimensional remote sensing attributes, the total number of spatial-temporal collocated MODIS-AERONET observations during a couple of years is relatively not large enough for estimation modeling. It leads to unstable feature selection subsets and drops the AOD estimation accuracy. In this paper, we propose a novel ensemble approach by aggregating multiple AOD estimators. Each estimator is modeled based on features selected from remote sensing attributes by using a subsampling strategy with instance perturbation. The ensemble approach provides aggregated retrievals of AOD with higher accuracy, while also providing an estimation of retrieval uncertainty. We conducted experiments to evaluate the empirical performance of the proposed approach on two years (2009-2011) of MODIS data over 197 global AERONET sites. The encouraging results clearly showed that aggregation of estimators modeled by multiple feature selection subsets leads to accuracy improvements and uncertainty reduction in AOD retrievals.

Keywords: *Feature Selection, Ensemble Regression, Aerosol Optical Depth, Instance Perturbation*

1. Introduction

MODIS (Moderate Resolution Imaging Spectroradiometer) is a key instrument aboard NASA's Earth Observing System (EOS) satellites Tera and Aqua. It is viewing the entire Earth's surface every 1 to 2 days, acquiring remote sensing data in 36 spectral bands ranging in wavelength from 0.4 μm to 14.4 μm. These massive streams of data will improve our understanding of global dynamics and processes occurring in the lower atmosphere over lands and oceans.

One of the most important tasks for MODIS is inferring the characteristics of aerosols. Aerosols are fine solid particles or liquid droplets in the air, including dust, fume, mist, smoke, fog, *et. al*,. Their concentration and chemical properties are important to study the Earth's radiation balance, climate change and environment protection. By using remote sensing observations and auxiliary parameters from MODIS, it is possible to estimate the degree to which aerosols prevent sun light passes through a column of atmosphere, a quantity parameter known as aerosol optical depth (AOD).

Traditionally in atmosphere study, AOD is retrieved according to NASA's deterministic forward-inversion approaches [1-4] by modeling atmosphere dynamics. However, many physical and chemical processes are involved in atmosphere dynamics and the complex nature of the Earth surface is very hard for construction of an accurate

ISSN: 2005-4254 IJSIP

retrieval model. By validation of MODIS AOD retrievals with Aerosol Robotic NETwork (AERONET) sites, previous studies concluded that MODIS retrievals have systematic bias in a global level [5-8]. For solving the problem, in recent years, several machine learning and data mining estimation methods have been actively developed and they explored data-driven regression models to achieve higher AOD retrieval accuracy. For example, D. J. Lary *et. al.*, applied neural networks and support vector machines to correct a persistent MODIS bias because of the dependency on surface types [9]. Arif Albayrak et al. utilized a neural network estimator to compensate against unknown sources of errors, nonlinearity in the datasets and the presence of non-normal distributions for improving AOD retrieval accuracy [10]. E. J. Hyer *et. al.*, developed empirical correction approaches based on surface boundary condition and regional microphysical bias to increase AOD estimation accuracy [11]. Zhuang Wang, *et al.*, treated the AOD retrievals as a multiple instance regression (MIS) problem due to different spatial resolution among collocated datasets. They approved that the instance pruning model in MIS is highly successful and can result in accurate estimations of AOD [12]. Vladan Radosavljevic, *et. al.*, proposed a continuous conditional random field model for regression of AOD. They provided strong evidence that the model can successfully exploit the inherent spatial-temporal properties of AOD data [13]. Slobodan Vucetic, *et. al.*, combined a neural network with a decision tree to analyze the conditions for improving AOD retrievals [14].

All these data-driven approaches constructed estimation models based on MODIS-AERONET collocated datasets, where the AOD measurements from AERONET sites work as prediction targets and the spatially and temporally collocated remote sensing observation attributes collected from MODIS act as driving attributes. In general, many potential informative observation attributes are constructed according to atmospheric domain knowledge or data analysis results. However, some of these constructed attributes might not be informative and make no contribution to prediction modeling. Meanwhile, the multiband satellite radiance observations contain high correlations among attributes. Building an estimation model from such redundant high-dimensional data breaks the assumption of independent and identically distributed random features in many data mining techniques and will result in the loss of estimation accuracy. In addition, compared with the number of these constructed high-dimensional satellite attributes, the total number of spatially and temporally collocated MODIS-AERONET records during a couple of years is relatively not large enough for estimation modeling. For example, there are no more than 3200 MODIS-AERONET synchronization records in one year over 200 global AERONET sites. Each record contains over 50 attributes, such as radiance in different wavelengths and auxiliary geometry parameters. Considering multiple atmospheric and surface conditions, the size of a general collocated dataset is not large enough for modeling the complex nature of AOD with so many attributes. Thereby, for dimension reduction and improving the estimation accuracy of AOD, a typical strategy is to perform feature selection before learning an estimation model.

In addition to dimension reduction for modeling and reducing redundancy from multiband radiance observations, there are several other benefits for feature selection in AOD estimation. Firstly, the selected features might be of interest to domain scientists focusing on identifying the most informative features for regression, so as to improve their certainty of measurement. Secondly, building an estimator from a small number of features could improve the estimator's generalization ability and reduce the risk of over-fitting. Thirdly, a small number of selected features could result in an easily interpretable estimation model.

Consequently, an important issue in AOD estimation is feature selection and picking out the informative driving features from all constructed attributes. With the constraint of limited collocation records, the feature subsets acquired from general feature selection

techniques are not stable with a few of instance perturbation. The unstable feature selection results decrease the AOD estimation accuracy.

Besides the point-estimation of AOD, another important issue in remote sensing of aerosols is the measurement of AOD retrieval uncertainty. We aim to identify a range where AOD retrievals will fall in with a high confidence level.

For tackling the above two issues, in this paper, we propose an ensemble estimation approach by aggregating multiple estimators of AOD. Each estimator is modeled based on feature selection attributes resulted from a subsampling strategy with instance perturbation effects. The key idea is explained as follows. Due to relative small ratio between the number of records and the number of features in the collocated dataset, a general feature selection method obtains an unstable feature subset, which leads to an unstable regression model achieving a local optimum with accuracy loss. We apply a subsampling strategy with instance perturbation to run a feature selection method for several rounds. Multiple feature subsets can be obtained and they are used to build multiple regression models. In this way, though several different regression models are constructed in different local optima, the ensemble estimator provides a better approximation of a true function by aggregating retrievals of AOD with higher accuracy. Meanwhile, the ensemble also can provide an estimation of retrieval uncertainty. We evaluated the proposed approach on MODIS collocated data over 197 global AERONET sites during April 2, 2009 and April 1, 2011. The experimental results clearly showed that aggregation of estimators based on multiple feature subsets leads to accuracy improvements and uncertainty reduction in AOD retrievals.

The rest of the paper is organized as follows. Section 2 illustrates the method of constructing an ensemble estimator. Section 3 validates the effectiveness of the proposed ensemble approach by experiments. Finally, Section 4 summarizes the paper and indicates future research directions.

2. Construction of an Ensemble Estimator

The MODIS-AERONET collocated points for a couple of years are a high-dimensional dataset with limited records. The mismatch of a large feature number and a relative small record number poses challenges for feature selection techniques and generally derives unstable selection attributes by instance perturbation. With a different resulted feature subset, an estimation algorithm may train a prediction model in a different local optimum. In this paper, we propose an ensemble learning approach. With it, a collection of single regression models are trained with different feature subsets. The ensemble estimations are obtained by averaging the outputs of these single models.

2.1. Feature Selection Techniques

There are multiple types of feature selection methods which are independent with an estimation model. The most common type of methods are ranking a feature according to its relevance to an estimation target, such as correlation criteria [15], mutual information [16], information gain [17], fisher score [18], *et. al.* These feature ranking methods have light computational costs, but ignore the redundant information among the top ranked features. Hence, it relies on practical experiments to decide the number of selected features at the top of a feature list. Meanwhile, some informative features may be filtered out since they are not at the top of ranks according to a specific criterion. The improved second type of methods aims to minimize global redundancy among selected features while maximizing their relevance to an estimation target [19-23]. They can discover a subset of more informative features and improve the overall prediction performance.

To test the generalization ability of our proposed ensemble approach, we select information gain and fisher score techniques from the first type of feature selection

methods, and pick the sparse multinomial logistic regression (SBMLR) from the second type of methods. All the three methods are widely used feature selection techniques.

2.1.1. Information Gain: This feature selection method is based on information theory. The information gain of a feature A reflects the randomness of categories after partitioning a dataset D by A. Specifically, the information gain of A is computed as,

$$Gain(A) = Info(D) - Info_A(D) \tag{1}$$

Where,

$$Info(D) = -\sum_{i=1}^{c} p_i \log_2(p_i) \tag{2}$$

$$Info_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j) \tag{3}$$

Here, Info(D) represents the expected information needed to classify a record in D, where c is the number of distinct category labels, p_i shows the probability of category C_i in D. $Info_A(D)$ computes how much more information is required for an exact classification after splitting D by a feature A, where v denotes the number of distinct values of A, $\frac{|D_j|}{|D|}$ acts as the ratio for the portion of records with feature A of value j in D.

2.1.2. Fisher Score: The fisher score method aims to find a subset of features, such that in the data space split by a selected feature, the distances between data points in different classes are as large as possible, while the distances between data points in the same class are as small as possible. In practice, we compute the fisher score for a feature A independently as below,

$$Fisher(A) = \frac{\sum_{k=1}^{c} n_k(\mu_k^A - \mu^A)^2}{\sum_{k=1}^{c} n_k(\sigma_k^A)^2} \tag{4}$$

Here, c is the number of classes. n_k denotes the number of k-th class in a dataset D. μ_k^A and σ_k^A represent the mean and standard deviation of k-th class, corresponding to the feature A. μ^A denotes the mean of the whole data set corresponding to the feature A.

2.1.3. Sparse Multinomial Longistic Regression (SBMLR): Given a dataset D with l records $\{(x^n, t^n)\}_{n=1}^{l}$, x^n is the n-th example vector with d attributes, t^n is the corresponding category label using 1 of c coding scheme. This feature selection method aims to minimize a penalized maximum-likelihood training expression as below,

$$M = E_D + \alpha E_w \tag{5}$$

Here,

$$E_D = \sum_{n=1}^{l} E_D^n = -\sum_{n=1}^{l}\sum_{i=1}^{c} t_i^n \log\{y_i^n\} \tag{6}$$

$$y_i^n = \frac{\exp\{a_i^n\}}{\sum_{j=1}^{c} \exp\{a_j^n\}}$$

Where

$$a_i^n = \sum_{j=1}^{d} w_{ij} x_j^n \tag{7}$$

$$E_w = \sum_{i=1}^{c} \sum_{j=1}^{d} | w_{ij} | \tag{8}$$

α is a regularization parameter balancing the bias-variance. w_{ij} is a weight of a linear model for the j-th attribute in the i-th example.

At a minimum of M, the partial derivatives of M with respect to the model parameters will be uniformly zero, deriving

$$| \frac{\partial E_D}{\partial w_{ij}} | = \alpha \text{ if } | w_{ij} | > 0 \text{ and } | \frac{\partial E_D}{\partial w_{ij}} | < \alpha \text{ if } | w_{ij} | = 0 \tag{9}$$

It means some weight parameters will be equal to zeros and thereby the corresponding features will be filtered out.

2.2. Measure Feature Selection Stability with Instance Perturbation

The MODIS-CALIPSO collocated dataset is relatively not large by comparing with its high-dimensional feature numbers. Therefore, by adopting different samples as a training set, we obtain different feature selection subsets. Suppose the total number of records in a training set is N, with a subsampling strategy, we randomly take Q=p• N samples (o<p<1, it is the percentage of N samples) as a training set. Consequently, feature selection is performed on each round of Q subsamples. Since a different feature selection round will compute a different feature weight, we measure the similarity between two rounds of feature selection results f_i, f_j according to their feature ranks. The feature with the least weight is assigned rank 1 and the best feature ranks d. By Spearman rank correlation coefficient, we compute the similarity between f_i, f_j as below,

$$Sim(f_i, f_j) = 1 - 6 * \frac{\sum_{k=1}^{d} (f_i^k - f_j^k)}{d * (d^2 - 1)} \tag{10}$$

Here, f_i^k is the rank for feature k in the i-th round.

The total stability of a feature selection method by instance perturbation for t rounds is measured as below,

$$Stability = \frac{2 * \sum_{i=1}^{t-1} \sum_{j=i+1}^{t} Sim(f_i, f_j)}{t * (t - 1)} \tag{11}$$

2.3. Construction of an Ensemble Estimator

Construction of an ensemble estimator consists of two steps. The first step involves training a set of different estimators by multiple running of a feature selector. Variation in

the feature selector can be achieved by instance level perturbation during training. The second step weighted averages the AOD retrieval results from these single estimators.

Assuming that by t rounds of instance perturbation, we have t different feature selection resulted subsets $\{f_1,f_2,\ldots,f_t\}$. With each of this subset, we can use a regression model M to obtain a single estimator $M(f_i)$. The ensemble estimator E is defined as below,

$$E = \sum_{i=1}^{t} w_i \times M(f_i) \tag{12}$$

Here, w_i is the weight to a regression model outputs $M(f_i)$. This can be used to accommodate for putting different significance on different models according to some domain criterion.

2.4. Regression Accuracy Measures

The accuracy of AOD regression is computed by three widely accepted measures: correlation coefficient (Corr), mean square error (MSE) , R^2. Their equations are listed as below. E_AOD denotes the AOT estimation of a regression model. A_AOD denotes AERONET AOD. N is the number of AOD retrievals. $\overline{E\_AOD}$, $\overline{A\_AOD}$ present the average of AOD estimations and AERONET AODs respectively. The std(A_AOD) denotes the standard derivation of AERONET AOD retrievals.

$$Corr = \frac{\sum_{i=1}^{N}(E\_AOD_i - \overline{E\_AOD}) \cdot (A\_AOD_i - \overline{A\_AOD})}{\sqrt{\sum_{i=1}^{N}(E\_AOD_i - \overline{E\_AOD})^2 \cdot \sum_{i=1}^{N}(A\_AOD_i - \overline{A\_AOD})^2}} \tag{13}$$

$$MSE = \frac{\sum_{i=1}^{N}(A\_AOD_i - E\_AOD_i)^2}{N} \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(A\_AOD_i - E\_AOD_i)^2}{\sum_{i=1}^{N}(A\_AOD_i - \overline{A\_AOD})^2} \tag{15}$$

3. Experimental Results

In this section, we present the results of an ensemble AOD estimator on the spatial-temporal collocated MODIS-AERONET dataset. Firstly, the synchronization data set is briefly described. Next, we analyze the different feature selection subsets by information gain, fisher score and SBMLR, and further show the similarity between feature selection subset pairs and analysis their stability correspondingly. Finally, we compare the ensemble estimator results with the results from single estimators.

3.1. MODIS-AERONET Collocated Data Sets

AERONET is a global observation network over multiple hundreds of sites providing ground-based aerosol retrievals. These retrievals are accurate for many research modeling and are widely accepted as the ground-truth for validation of satellite AOD retrievals [1-

7]. In our experiments, Level 2.0 cloud-screened and quality-assured AERONET data over 197 global observation sites from April 2, 2009 to April 1, 2011 are collected. These sites cover different surface types, such as land, coast, desert and marine.

MODIS is an important aerosol retrieval instrument aboard the satellite Tera and Aqua and views the Earth's surface in 36 spectral bands. In the experiments, we used MODIS/Aqua Collection 005 product suites between April 2, 2009 and April 1, 2011. It contains three MODIS products: level 2 aerosol product MYD04_L2 at a 10-km resolution, calibrated radiance data MYD02SSH at a 5-km resolution and cloud mask product MYD35 in a resolution of 1-km. These datasets are spatial-temporally synchronized in the spatial coincidence square region of size 40km×40km surrounding an AERONET site.

MODIS and AERONET datasets are collocated by the following spatial-temporal coincidence criteria proposed by Ichoku *et al.,* (2002): spatially, MODIS remote sensing attributes are averaged in a 40km× 40km rectangle with an AERONET site at the center; temporally, AERONET observations are averaged within ± 30 minutes of MODIS overpass.

During the two years of study period from April, 2009, we collected 6351 collocated records covering 197 AERONET sites globally in the MODIS-AERONET sync data. Half of the records act for feature selection and model training, and the other half works as a test dataset. Both the training and test dataset contain 53 driving attributes. The attribute details are listed in Table 1. We can observe that some of these attributes might provide redundant information for regression modeling, such as the seven wavelength radiance observations for a record. Meanwhile, considering 53 attributes, no more than three thousand and two hundred records in a training dataset might not be large enough for a stable feature selection and regression modeling. Thereby, feature selection is necessary for improving the regression accuracy of AOD retrievals.

Table 1. Driving Attributes Constructed at 40km×40km Resolution

Index	Attribute Explanations
1-14	Means and standard derivations of 7 radiance measurements related to aerosol retrievals
15-20	Means and standard derivations of surface reflectance at 3 wavelengths
21-24	Solar zenith, Solar azimuth, Sensor zenith, Sensor azimuth
25	Scattering angle
26-27	Mean and standard derivation of angstrom exponent
28-30	NDVI_swir, NDVI_swir2 and standard derivation of NDVI_swir
31-32	Mean and standard derivation of cloud fractions
33-37	Percentage over water, costal, desert, land and land_only_flag
38	The number of pixels without clouds
39	Aerosol types provided by MODIS
40-42	AERONET site altitude, **MODIS** surface altitude, standard derivation of MODIS surface altitudes
43-45	The distances from three control clustering centers
46-47	Mean and standard derivation of MODIS AOD retrievals
48-49	Latitude and Longitude
50-53	Year, month, day, hour

3.2. Feature Selection Stability

In the training set, we have 3175 records in total. Each record contains 53 attributes. For feature selection, the parameter p is set to 90% and t is set to 10. Thereby, we

randomly select 90%*3175=2858 records for feature selection in each round and we repeat the same feature selection method for 10 rounds. For the method of Information Gain, the Table 2 reports the spearman rank correlation coefficients between any two rounds of feature selection results. The first row and the first column list the index of a feature selection round. The similarity values range from 0.173 to 0.752. It shows that the information gain unstably selects features by instance perturbation. The overall stability computed by Equation (11) is 0.490.

Table 2. Similarity between Features Selected by Information Gain

#	2	3	4	5	6	7	8	9	10
1	0.495	0.560	0.173	0.748	0.237	0.333	0.612	0.464	0.386
2		0.578	0.365	0.691	0.620	0.618	0.662	0.739	0.198
3			0.264	0.443	0.525	0.553	0.425	0.679	0.572
4				0.257	0.588	0.631	0.486	0.365	0.231
5					0.390	0.470	0.596	0.716	0.226
6						0.752	0.601	0.595	0.231
7							0.597	0.710	0.430
8								0.537	0.321
9									0.368

For fisher score, the similarities between selected features are presented in Table 3. Their values range from 0.044 to 0.788. The overall stability is 0.443. Compared with information gain, fisher score obtains a little less stable feature selection results. Table 4 reports the similarities between features selected by SBMLR. The minimum value is 0.033 and the maximum is 0.538. The overall stability is 0.3234. Both the maximum and the overall stability values are far less than those of information gain and fisher score. It can be understood that SBMLR minimize the global redundancy among all features and the complex nature and the size of a training set with high-dimensional features make SBMLR lead to less stable feature selection results.

Table 3. Similarity between Features Selected by Fisher Score

#	2	3	4	5	6	7	8	9	10
1	0.396	0.476	0.430	0.304	0.161	0.408	0.551	0.573	0.336
2		0.542	0.636	0.521	0.500	0.606	0.386	0.590	0.483
3			0.367	0.579	0.555	0.507	0.270	0.170	0.494
4				0.322	0.446	0.460	0.500	0.427	0.413
5					0.788	0.597	0.246	0.375	0.523
6						0.482	0.044	0.208	0.399
7							0.396	0.523	0.543
8								0.576	0.438
9									0.396

3.3. Construction of an Ensemble Estimator

We apply a neural network (ANN) with one hidden layer as the regression model. The inputs to an ANN are the selected features by Information Gain, Fisher Score or SBMLR. The output is AOD. By practical experiments, the number of neurons in the hidden layer is set to 6. For each of the three feature selection techniques, we run it for 10 times on the 90% of a training dataset by instance perturbation. Thereby, we have 10 different feature selection subsets. Based on each of the feature subset, we training an ANN model for 20 times. The averages and standard derivations of the 10 models for 20 times are reported in

Table 5. In the corresponding ensemble estimator, 10 models are averaged with weight w_i setting to 0.1. The means and standard derivations of an ensemble estimator for 20 times are also listed in Table 5. By comparison of three accuracy measurements, we see clearly that an ensemble estimator performs significantly more accurate than a single estimator.

Specifically, the standard derivation for estimations on each point by 10 models provides an indicator of retrieval uncertainty. Though we apply three different feature selection methods, their estimation standard derivation falls in the similar range. It suggests the uncertainty level by an ANN model based on the selected features. Compared with the standard derivation of accuracy measurements for single estimators, the uncertainty of an ensemble estimator has significant decrease.

Table 4. Similarity between Features Selected by SBMLR

#	2	3	4	5	6	7	8	9	10
1	0.270	0.209	0.431	0.172	0.088	0.305	0.172	0.271	0.087
2		0.291	0.033	0.349	0.316	0.483	0.446	0.290	0.292
3			0.326	0.465	0.341	0.055	0.447	0.457	0.370
4				0.251	0.524	0.249	0.126	0.452	0.368
5					0.405	0.234	0.538	0.321	0.507
6						0.294	0.353	0.353	0.455
7							0.385	0.398	0.250
8								0.324	0.478
9									0.325

Table 5. AOD Regression Accuracy Comparison between a Single Estimator and an Ensemble Estimator

Method	Measure	Single	Ensemble
Information Gain	Corr	0.8967±0.0025	0.9003±0.0020
	MSE	0.0042±0.0001	0.0040±0.0000
	R^2	0.8036±0.0047	0.8104±0.0039
Fisher Score	Corr	0.8955±0.0020	0.8990±0.0018
	MSE	0.0044±0.0001	0.0042±0.0000
	R^2	0.8016±0.0036	0.8080±0.0030
SBMLR	Corr	0.8959±0.0020	0.8993±0.0020
	MSE	0.0042±0.0001	0.0041±0.0000
	R^2	0.8022±0.0038	0.8087±0.0037

4. Conclusions and Future Work

The spatial-temporally collocated records between a satellite facility and AERONET provide an excellent dataset for construction of a data mining regression model for AOD. However, the remote sensing data collects many attribute observations. With such size of collocated records and feature numbers, a feature selection method produces unstable results and it leads to the loss of AOD regression accuracy. In this paper, we apply instance perturbation on the same feature selection method for several rounds and obtain multiple feature selection subsets. Next, we aggregate the regression estimators based on the different subsets together and make an ensemble estimator. We test the proposed approach on the MODIS-AERONET collocated dataset over 197 global sites for 2 years. Experimental results show that an ensemble estimator achieves significantly more accurate AOD retrievals than a single estimator. Meanwhile, the standard derivation of the ensemble estimator suggests that it is more robust than these single estimators as well.

The ensemble estimator technique might open many new avenues for further research. For example, we have applied instance perturbation to obtain several feature selectors in this paper. Further we will explore multiple other approaches to make variations in the feature selectors, such as different feature selection techniques, feature level sensitivity perturbation, etc. The measuring of robustness for each feature selection technique and regression approach is also an important task in our next research.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 61272272 and U1531122, and by the Natural Science Foundation of Hubei province under Grant 2015CFA058.

References

[1] L. A. Remer, Y. J. Kaufman and D. Tanré, "The MODIS aerosol algorithm, products and validation", Journal of Atmospheric Sciences, vol. 62, **(2005)**, pp. 947-973.
[2] R. C. Levy, L. A. Remer and S. Mattoo, "Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance", Journal of Geophysical Research, D13211, doi: 10.1029/2006JD007811, vol. 112, **(2007)**.
[3] R. C. Levy, L. A. Remer and O. Dubovik, "Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land", Journal of Geophysical Research, D13210, doi: 10.1029/2006JD007815, vol. 112, **(2007)**.
[4] R. C. Levy, S. Mattoo and L. A. Munchak, "The Collection 6 MODIS aerosol products over land and ocean", Atmospheric Measurement Techniques, vol. 6, no. 11, **(2013)**, pp. 2989-3034.
[5] D. A. Chu, Y. J. Kaufman and C. Ichoku, "Validation of MODIS aerosol optical depth retrieval over land", Geophysical Research Letters, doi: 10.1029/2001GL01320, vol. 29, no. 12, **(2002)**.
[6] M. A. Friedl, D. S. Menashe and B. Tan, "MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets", Remote Sensing of Environment, vol. 114, no. 1, **(2010)**, pp. 168-182.
[7] R. C. Levy, L. A. Remer and R. G. Kleidman, "Global evaluation of the Collection 5 MODIS dark-target aerosol products over land", Atmospheric Chemistry and Physics, vol. 10, no. 21, **(2010)**, pp. 10399-10420.
[8] X. Xia, "Significant overestimation of global aerosol optical thickness by MODIS over land", Chinese Science Bulletin, vol. 51, no. 23, **(2006)**, pp. 2905-2912.
[9] D. J. Lary, L. A. Remer and D. MacNeill, "Machine learning and bias correction of MODIS aerosol optical depth", IEEE Geoscience and Remote Sensing Letters, vol. 6, no. 4, **(2009)**, pp. 694-698.
[10] A. Albayrak, J. Wei and M. Petrenko, "Global bias adjustment for MODIS aerosol optical thickness using neural network", Journal of Applied Remote Sensing, vol. 7, no. 1, **(2013)**, pp. 073514-073514.
[11] E. J. Hyer, J. S. Reid and J. Zhang, "An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals", Atmospheric Measurement Techniques, vol. 4, no. 3, **(2011)**, pp. 379-408.
[12] Z. Wang, V. Radosavljevic and B. Han, "Aerosol Optical Depth Prediction from Satellite Observations by Multiple Instance Regression", SDM, **(2008)**, pp. 165-176.
[13] V. Radosavljevic, S. Vucetic and Z. Obradovic, "Continuous Conditional Random Fields for Regression in Remote Sensing", ECAI, **(2010)**, pp. 809-814.
[14] S. Vucetic, B. Han and W. Mi, "A data-mining approach for the validation of aerosol retrievals", IEEE Geoscience and Remote Sensing Letters, vol. 5, no. 1, **(2008)**, pp. 113-117.
[15] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection", Journal of Machine Learning Research, vol. 3, **(2003)**, pp. 1157–1182.
[16] C. Lazar, J. Taminau and S. Meganck, "A survey on filter techniques for feature selection in gene expression microarrayanalysis", IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 9, **(2012)**.
[17] J. Han, M. Kamber and J. Pei, "Data Mining Concepts and Techniques", Elsevier, **(2011)**.
[18] P. E. H. R. O. Duda and D. G. Stock, "Pattern Classication", Wiley-Interscience Publication, **(2001)**.
[19] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data", Journal of bioinformatics and computational biology, vol. 3, no. 02, **(2005)**, pp. 185-205.
[20] A. Unler, A. Murat and R. B. Chinnam, "PSO: a maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification", Information Sciences, vol. 181, no. 20, **(2011)**, pp. 4625-4641.
[21] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy", Journal of Machine Learning Research, vol. 5, **(2004)**, pp. 1205-1224.

[22] F. C. Ding and H. L. Peng, "Feature selection based on mutual information: Criteria of maxdependency, max-relevance, and min-redundancy", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, **(2005)**, pp. 1226-1238.
[23] G. C. Cawley, N. L. C. Talbot and M. Girolami, "Sparse multinomial logistic regression via bayesian l1 regularisation", In Advances in Neural Information Processing Systems, **(2007)**.

Author

Bo Han, received the B.S. degree in computer science and its applications from Wuhan University of Technology, Wuhan, China in 1993, and the M.S. degree in computer science from Wuhan University, Wuhan, China in 1996, and the Ph.D. degree in computer and information science from Temple University, Philadelphia, USA in 2007.

He is currently an Associate Professor with the International School of Software in Wuhan University, Wuhan, China. His research interests are in spatio-temporal data mining, text mining and machine learning.