Accuracy improvement and Blending of MHS, SSMIS, and GMI rain rate product using DNN and EBMA

Kwangjin Kim (1),Yang-Won Lee (1)

[1] Pukyong Nat. Univ., 48513 45, Yongso-ro, Nam-gu, Busan, Korea
Email: somakkj@gmail.com; modconfi@pknu.ac.kr;

KEY WORDS: Precipitation, Blending, Ensemble Bayesian model averaging(EBMA), Deep neural network

ABSTRACT: In this study, deep neural network is utilized as another approach for improving accuracy of the precipitation based on microwave-sensor. And The ensemble Bayesian model averaging(EBMA), which employs a weighting scheme for each member using posterior probability, in order to produce a more improved blending of precipitation from multi-satellite and to evaluate the effect of accuracy improvement. Experiments to improve rain rate were carried out based on data obtained from Global Precipitation Measurement (GPM) Microwave Imager (GMI). Input data for the DNN model include 7 brightness temperatures (Tb), ice water path (IWP), convective rain rate, scattering index (SI) and land sea mask is used. The experiment for blending of precipitation product was performed using rain rate product of three satellites and sensors, namely GMI of GPM core observatory, special sensor microwave imager/sounder (SSMI/S) of the Defense Meteorological Satellite Program (DMSP) F16 and microwave humidity sounder (MHS) of NOAA-18. In both experiments, precipitation product of the Dual-frequency Precipitation Radar (DPR) of CO was used as reference data. The probability density function(PDF) of gamma distribution combined with logistic regression is used to estimate the probability and quantity of precipitation for considering the characteristics of precipitation. And then, the exponent for these two functions and the percentile threshold of the cumulative density function were set by optimizing simulations. After that, the validation statistics of the blending precipitation through comparison with precipitation obtained from DPR is carried out.

1. INTRODUCTION

Precipitation is an important factor in the hydrological cycle, and global distribution and intensity are essential for understanding and feedback of the Earth system. The estimate by remote sensing is used for measurement of the global precipitation. Satellite remote sensing using microwaves is easy to determine the global precipitation distribution and its amount. Passive microwave(PMW) sensors are generally used for estimating the precipitation. Rain rate product of The Goddard profiling(GPROF) algorithm provided from the global precipitation measurement(GPM) mission was improved compared with previous estimating methods. However, improvements are still needed. This is because it is difficult to express the microscopic characteristics of the state of water droplets in the clouds and the precipitation of the ground in a linear relationship. Artificial neural network(ANN) has the advantage of being able to express these nonlinear relationships(Sanò et al. 2015). Recently, deep neural network(DNN) that solves the disadvantages of the existing ANN has been used in various fields. Precipitation estimation using satellites requires not only accuracy but also large spatial coverage. It is difficult to identify the global precipitation distribution using a single satellite. Therefore, various institutions generate and distribute synthetic fields. Most of them use simple merging or the closest estimates for overlapping regions. This means that no consideration is given to differences in individual uncertainties for each dataset. EBMA is a weighted average of post-probability density functions using post-probability as a weight. It is possible to generate predictions by considering the difference of each uncertainty of the individual data for the overlapping regions when generating the composite field. In this study, DNN was used to improve the accuracy of rainfall intensity product based on data generated by GPROF algorithm. In addition, we confirmed the applicability and improvement in the overlapping areas by using EBMA for the blending product.

2. DATA

2.1 Data for DNN test

For the DNN test to improve the accuracy of the precipitation product based on the PMW sensor, seven Tb data, IWP, convective rain rate, SI, and land sea mask data were used as input data. Tb data was obtained from Global precipitation mission(GMI) core observatory(CO) / GPM microwave imager(GMI) level 1C data provided by PPS. The IWP and convective rain rate are similarly provided by PPS and Level 2A, which is generated by the GPROF algorithm. SI was obtained using the SI calculation formula of the defense meteorological satellite program(DMSP) / special sensor microwave imager/sounder(SSMI/S). Land sea mask was used to obtain reanalysis data of European centre for medium-range weather forecasts(ECMWF).

2.2 Data for EBMA test

For the blending test, the rain rate products provided by PPS were obtained. GPM CO / GMI, DMSP F-16 / SSMIS, and national oceanic and atmospheric administration(NOAA)-18 / Microwave Humidity Sounder(MHS) GPROF rain rate product Level 2A data were used.

2.3 Reference data

Dual-frequency precipitation radar(DPR) is one of the prime instruments onboard the GPM CO. It uses an active microwave sensor, that is radar. This allows more direct information acquisition than PMW. Therefore, DPR rainfall data was used as a reference for improving precipitation product and calculating weight for blending. In this study, the surface precipitation data of the GPM Ku Level 2 normal scan(NS) region was used to construct many matchups.

2.4 Study area and period and preprocessing of data

The study area is set in the region of Northeast Asia, which is not verified much compared to the existing studies, and is set from 20˚N to 50˚N degrees and from 100˚E to 150˚E. Data were ingested from 2016 to 2018 in the test for DNN t, and data from April 2014 to April 2016 were used for the composite field test. All inputs carried out resampling process with same grid for match-up processing. The grid was resampled with 0.1 degree for DNN applications and 0.25 degree for blending tests.

3. METHODS

3.1 Deep neural network

Artificial intelligence is a technology that simulates human intelligence through computer training. Artificial neural networks are a kind of artificial intelligence such as machine learning and are widely used for classification and prediction in various studies. Artificial neural networks have the advantage of estimating complex nonlinear functions (Chen et al., 2006;, 2009). Deep neural networks combine the advantages of neural networks and machine learning through deep networks of multiple hidden layers. Previous multilayer neural networks had problems with local minima. Local minima means that to stop the change of weights at local convergence values before the entire data is represented. Since this is a local convergence value, it is difficult to express the relationship of the overall nonlinear model to be expressed.In the DNN, the local minima problem is improved with the activation function that prevents the gradient loss of the loss function during the backpropagation process, which updates the weight and bias set in the forward and reverse directions of the network. General machine learning techniques also present the problem of overfitting. Overfitting means lower errors in the training data due to excessive training but larger errors in the prediction. Overfitting can be prevented through regularization techniques such as L1, L2, and dropout techniques. The DNN needs to be optimized through the above techniques to represent the nonlinear relationship between the input and output data. In this study, an optimal DNN model was constructed to represent the relationship between input data and precipitation through tests on hidden layer setup, dropout, and L2 regularization.

3.2 Ensemble Bayesian model averaging(EBMA)

EBMA is a statistical ensemble technique that uses the PDF as a weighted average of individual members (Raftery et al., 2005; Sloughter et al., 2007). In terms of the value of the second data, the EBMA PDF is expressed as (1).

$$\mathrm{p}(\mathrm{y}|f_1,\dots,f_K) = \sum_{k=1}^{K} w_k h_k(y|f_k) \tag{1}$$

where $w_k$denote the posterior probability as the weight, $h_k(y|f_k)$ is a posterior PDF. Precipitation is a type of gamma distribution with a high frequency of zero and a significant skewness of the distribution. In order to apply EBMA to precipitation data, logistic regression and gamma distribution were used. As shown in (2), it is expressed as the sum of non-precipitation ($I[y=0]$) and precipitation ($I[y>0]$) elements.

$$h_k(y|f_k) = p(y=0|f_k)\, I[y=0] + p(y>0|f_k) g_k(y|f_k) I[y>0] \tag{2}$$

The non-precipitation component is the form of logistic regression. It can be represented by (3).

$$p(y=0|f_k) \equiv \log\frac{p(y=0|f_k)}{p(y>0|f_k)} = a_{0k} + a_{1k}f_k^m + a_{2k}\delta_k \tag{3}$$

Where, subscripts a are coefficient of logistic regression. And, δ is the adjustment fator, which is 1 when the member is zero, and 0 otherwise. Also m is the number of root converting to prediction. The precipitation components can be represented using a gamma distribution. The gamma distribution is expressed as PDF as (4) with a shape parameter (α) and a scale parameter (β).

$$g_k(\mathrm{y}|f_k) = \frac{1}{\beta_k^{\alpha_k}\Gamma(\alpha_k)} y^{\alpha_k - 1}\exp(-y/\beta_k) \tag{4}$$

The gamma distribution parameters for each data can be obtained using (5) and (6) using the mean(μ_k) and variance(σ_k^2) of the data values.

$$\mu_k = b_0 + b_1 f_k^m = \alpha_k\beta_k \tag{5}$$

$$\sigma_k^2 = c_0 + d_1 f_k = \alpha_k\beta_k^2 \tag{6}$$

Thus, the final EBMA model is the same as (7).

$$\mathrm{p}(\mathrm{y}|f_1,\ldots,f_K) = \sum_{k=1}^{K} w_k\,[p(y=0|f_k)\,I[y=0] + p(y>0|f_k)g_k(y|f_k)I[y>0]] \tag{7}$$

4. RESULT

In order to improve the accuracy of rain rate data, the training and optimizing the DNN model of two stage were carried out with 70% of the total matchups. These two steps are models for precipitation classification and estimating regression using DNN. As a result of the optimization through iteration, for the classification, the hidden layer is consisted of two layers with 200 nodes and the epoch was set to 100. And the hidden layer is consisted of three layers with 100 nodes, and the epoch was set to 240 for regulation. Regularization and dropout experiments were carried out only on the DNN regression model. For regularization, it is confirmed that the accuracy of the model decreases as the dropout and L2 regularization ratio increases. This means that the training data contains the variety of information, the training dataset can fully satisfy the ability of regularization. Table 1 shows the error rates of the DNN classification model. Compared to GPROF, probability of detection (POD) and Heidke skill score (HSS) were improved. False alarm ratio (FAR) was somewhat higher is was slight difference.

Table 1 Probability of detection(POD), false alarm rate(FAR) and Heidke skill score (HSS) of DNN classification and GPROF.

	POD	FAR	HSS
GPROF	0.818	0.084	0.743
DNN	0.912	0.097	0.814

When comparing the rain rate data obtained from the GPROF data and the DNN, the correlation coefficient improved by 17.0% ((0.778-0.665) /0.665) in the verification data. The root mean square error(RMSE) also showed improvement of 25.0% ((1.918-1.439) /1.918) in the verification data.

Table 2 . Correlation coefficient(r), mean bias(MB), mean absolute error(MAE) and root mean square error(RMSE) of rain rate retrieved from GPROF and DNN according to rain rate of DPR for training and validation. DNN_only_Tb is the model generated by only brightness temperature and DNN is generated by all input variable.

		r	MB	MAE	RMSE
Train	GPROF	0.662	0.075	0.964	1.932
	DNN_only_Tb	0.743	0.016	0.785	1.527
	DNN	0.785	0.002	0.710	1.414
Validation	GPROF	0.665	0.083	0.963	1.918
	DNN_only_Tb	0.744	0.013	0.787	1.529
	DNN	0.778	0.006	0.718	1.439

In the experiment for generating the blending data, the exponent , m, for (3) and (5) were set to 1/2 through iteration. Since the rain rate cannot be directly obtained from the EBMA PDF, the rain rate must be derived by converting the PDF into a cumulative distribution function and taking the inverse function of this CDF. As a result of selecting and comparing various threshold on the CDF, it was finally set to 70%. Table 2 shows the error from EBMA data by comparing mean and median ensembles, a commonly used method of blending.

Table 3 The correlation and error statistics of mean ensemble, median ensemble and EBMA blending.

Method	**r**	**Mean bias (mm/hr)**	**MAE (mm/hr)**	**RMSE (mm/hr)**
EBMA	0.694	-0.081	0.770	1.352
Ensemble mean	0.623	-0.169	0.825	1.568
Ensemble median	0.634	-0.244	0.807	1.523

5. CONCLUSION

The two DNN models were used to improve the precipitation product of GPROF, which is rain rate estimation algorithm using PMW sensor. In order to generate the optimal DNN model, the optimal hidden layer, the number of nodes, and the training frequency epoch were selected through an iterative process. So classification of precipitation using the DNN showed higher score on POD and HSS than GPROF. Rain rate improved by the DNN model showed a higher and more consistent correlation coefficient than the output of the existing GPROF algorithm. The monthly comparison also showed the consistent error rate. Blending experiment of multiple precipitation data was done using EBMA. Compared with the average and median ensembles that are generally used, it showed highest correlation coefficient and lowest error. In addition, when the distribution was confirmed through the mapping of the results, the discontinuous distribution of overlapping regions was not seen, and the accuracy improvement effect was shown compared to the ensemble members.

REFFERENCE

Sanò, P., Panegrossi, G., Casella, D., Di Paola, F., Milani, L., Mugnai, A., Petracca, M., and Dietrich, S., “The Passive microwave Neural network Precipitation Retrieval (PNPR) algorithm for AMSU/MHS observations: description and application to European case studies,” Atmospheric Measurement Techniques, 8(2), 837-857 (2015).

Chen, Y., Aires, F., Francis, J. A., and Miller, J. R., “Observed relationships between Arctic longwave cloud forcing and cloud parameters using a neural network,” Journal of climate, 19(16), 4087-4104 (2006).

Marzban, C. [Basic statistics and basic AI: neural networks. In Artificial Intelligence Methods in the Environmental Sciences], Springer, Dordrecht, 15-47 (2009).

Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M., “Using Bayesian model averaging to calibrate forecast ensembles,” Monthly Weather Review, 133(5), 1155-1174 (2005).

Sloughter, J. M. L., Raftery, A. E., Gneiting, T. and Fraley, C., “Probabilistic quantitative precipitation forecasting using Bayesian model averaging,” Monthly Weather Review, 135(9), 3209-3220 (2007)