

Black-Box Classifier Interpretation Using Decision Tree and Fuzzy Logic-Based Classifier Implementation

Hansoo Lee and Sungshin Kim

Department of Electrical and Computer Engineering, Pusan National University, Busan, Korea



Abstract

Black-box classifiers, such as artificial neural network and support vector machine, are a popular classifier because of its remarkable performance. They are applied in various fields such as inductive inferences, classifications, or regressions. However, by its characteristics, they cannot provide appropriate explanations how the classification results are derived. Therefore, there are plenty of actively discussed researches about interpreting trained black-box classifiers. In this paper, we propose a method to make a fuzzy logic-based classifier using extracted rules from the artificial neural network and support vector machine in order to interpret internal structures. As an object of classification, an anomalous propagation echo is selected which occurs frequently in radar data and becomes the problem in a precipitation estimation process. After applying a clustering method, learning dataset is generated from clusters. Using the learning dataset, artificial neural network and support vector machine are implemented. After that, decision trees for each classifier are generated. And they are used to implement simplified fuzzy logic-based classifiers by rule extraction and input selection. Finally, we can verify and compare performances. With actual occurrence cased of the anomalous propagation echo, we can determine the inner structures of the black-box classifiers.

Keywords: Black-box models, Decision tree, Fuzzy inference system, Interpretation, Rule extraction

1. Introduction

The successful applications of black-box models in various fields such as engineering, science, marketing and medicine have provided an evidence of the black-box models over past years [1]. Among the black-box models, artificial neural networks (ANNs) and support vector machines (SVMs) are representative models and they have been proven their usefulness with remarkable performances. The artificial neural networks, which are inspired by microscopic biological models, have been widely used as solving many regression and classification problems [2, 3]. The support vector machines, which are based on the idea of structural risk minimization instead of empirical risk minimization, has been introduced as power equipments for solving regression and classification problems [4, 5].

However, as always, everything has two sides, and ANNs and SVMs are not an exemption. Even the ANNs and the SVMs attain remarkable performances in regression and classification problems, these models cannot provide delineation in a comprehensible form the process through which a given output generated by the models have been reached [1, 6]. It leads to

Received: Dec. 7, 2015
Revised : Mar. 23, 2016
Accepted: Mar. 24, 2016

Correspondence to: Sungshin Kim
(sskim@pusan.ac.kr)
©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

prevent to apply the black-box models in some problems which is essential that a system user should be able to validate the output of the models under all possible input conditions such as airlines, power plants, medical care, and so on. Therefore, researches and techniques for rule extraction from ANNs and SVMs have been introduced to improve this problem and aid in elucidation of their decisions [1, 6].

Rule extraction is the process of developing natural language-like syntax that describes the behavior of black-box models, and changes the models into a white-box systems by translating the internal knowledge of the models into a set of symbolic rules [7]. Many rule extraction algorithms have been designed to reveal the information concealed in the black-box models. In case of the rule extraction algorithms for the ANNs, there are NeuroRule [8], RX [9], GLARE [10], OSRE [11], etc. And in case of the rule extraction algorithms for the SVMs, there are SQREx-SVM [12], RulExSVM [13], SVM_DT [14], and so on.

In this paper, we propose a method to implement a fuzzy logic-based classifier implementation using a decision tree from the black-box classifiers. First, we choose a decision tree-based indirect interpretive method in order to delieate the black-box models. Second, we convert the derived decision trees to fuzzy logic-based classifiers for providing a linguistic variable-based classificaton method to users which is easily comprehensible. We choose both artificial neural network and support vector machine. Further, an anomalous propagation echo is selected in this paper as an object of classification which occurs frequently in radar data and becomes the problem in a precipitation estimation process.

The rest of the paper is organized as follows. In Section 2, we explain the entire proposed system which consists of ANN, SVM, decision tree, rule extraction, and fuzzy logic-based classifier. And in Section 3, we describe anomalous propagation echo as a target of classification methods. After that, the experimental results with actual anomalous propagation echo occurrence cases are presented in Section 4. Finally, conclusion and future work are shown in Section 5.

2. Interpretation of Black-Box Classifiers

2.1 ANN and SVM

An ANN is a biologically inspired computational model formed from lots of artificial neurons, connected with weights which constitute the neural structure. Although a single neuron can perform certain simple information processing functions, known as a perceptron model, the power of neural computations comes

from connecting neurons in a network. ANNs are capable of processing extensive amounts of data, and making accurate regression and classification results [3].

A SVM is a binary classification method that divides the given data into two groups in the best possible way by using hyperplanes. This method is based on a structural risk minimization method to reduce the error rather than the empirical risk minimization method used in traditional statistical learning theory. In other words, after the division of an entire group into subgroups, a decision function is selected. This function can minimize the empirical risk for the subgroups. Thus, the SVM method has the advantage of achieving great performance in classification, regression, and estimation processes by using a relatively low amount of the given learning data. Without any knowledge of the mapping, the SVM finds the optimal hyperplane by using the dot product functions in feature space using kernel functions. The solution of the optimal hyperplane can be written as a combination of a few input points that are called support vectors [4, 5].

The similarities between ANN and SVM is as follows. First, both ANN and SVM produce black-box model, which was the main motivation behind the rule extraction studies. Second, they can deal with nonlinear models using their own properties. In case of ANN, hidden layers and nonlinear activation functions make the ANN handle nonlinear systems. And in case of SVM, mapping to higher dimensional feature space and nonlinear kernel function make the SVM handle nonlinear systems. Further, their decision function form looks similar as following equations.

$$f(x) = \text{sign} \left(\sum_{i=1}^{lhs} \omega_i Z_i(x) \right), \quad (1)$$

$$f(x) = \text{sign} \left(\sum_{s=1}^{sv} \alpha_s y_s K(\chi_s, \chi) + b \right), \quad (2)$$

where lhs is last hidden layer, ω_i is the weight from last hidden layer to the output layer in equation (1), and where sv is the model support vectors, α_s is Lagrange multiplier, $K(\chi_s, \chi)$ is a kernel function in equation (2), respectively.

2.2 Rule Extraction Using Decision Tree and Fuzzy Logic-Based Classifier Implementation

Rule extraction methods search through inner structure of given classifiers and analyze their operating principle. It is important to figure out how the classifier derives results using given data especially the case which both derivation process and result are

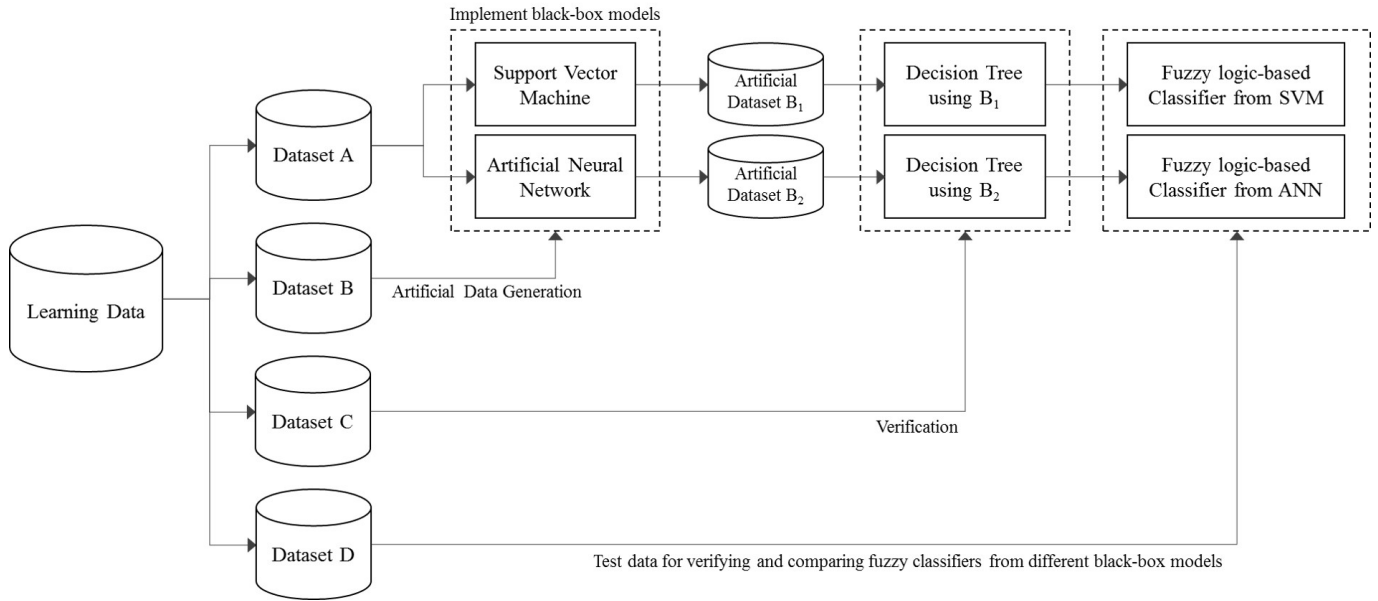


Figure 1. Entire process to establish fuzzy inference system from support vector machine.

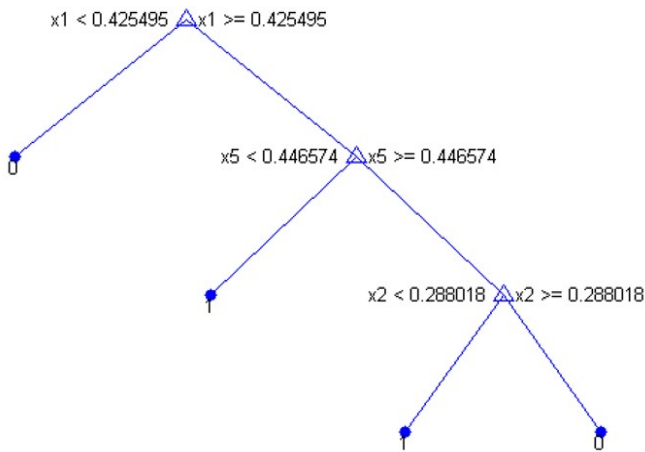


Figure 2. An actual implementation of decision tree from artificial neural network.

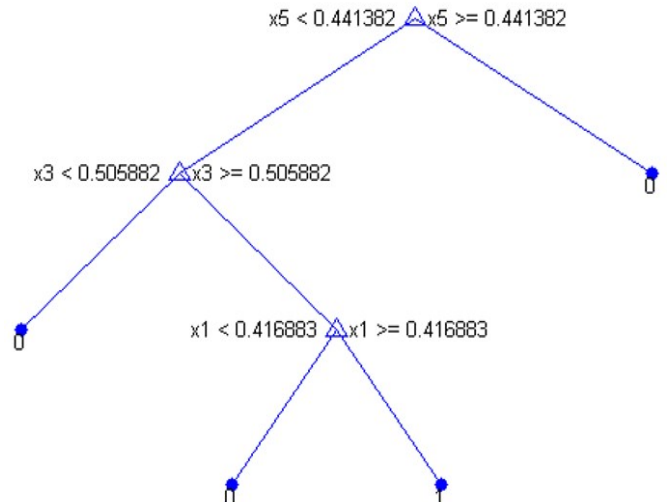


Figure 3. An actual implementation of decision tree from support vector machine.

crucial such as power plant, airlines, medical care, and so on. There are lots of ongoing researches about analyzing the ANNs and SVMs. In case of the rule extraction algorithms for the ANNs, there are NeuroRule [8], RX [9], GLARE [10], OSRE [11], etc. And in case of the rule extraction algorithms for the SVMs, there are SQREx-SVM [12], RulExSVM [13], SVM_DT [14], and so on.

Among these rule extraction methods, we select a method which uses a decision tree. This method is originated from the rule extraction method for SVM [15] using decision tree and artificially labelled dataset. The idea for creating artificially

labelled datasets where the given class is replaced by the classified results by the SVM and ANN without even looking at their individual inner structures. The artificially generated dataset can be collaborated with another classification methods which are comprehensible their entire principles.

The rule extraction method using decision tree consists of the following steps. The steps are summarized in Figure 1.

- Step 1: Normalize given learning data.
- Step 2: Divide the normalized data into four pieces, and

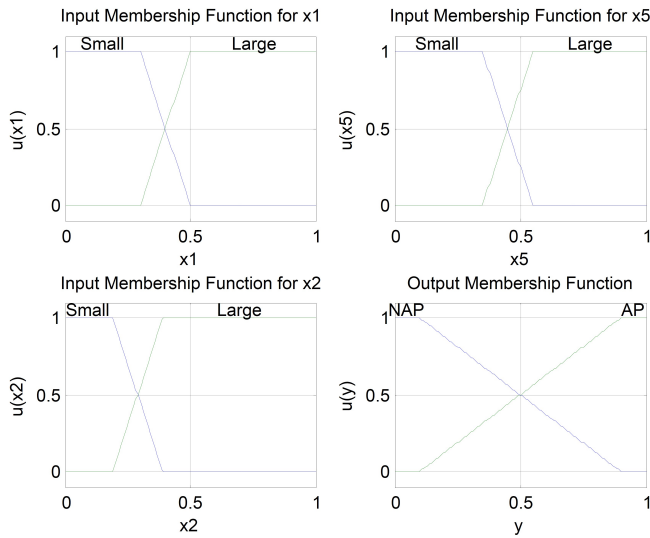


Figure 4. Input and output membership functions of fuzzy logic-based classifier from artificial neural network. AP, anomalous propagation echo; NAP, non-anomalous propagation echo.

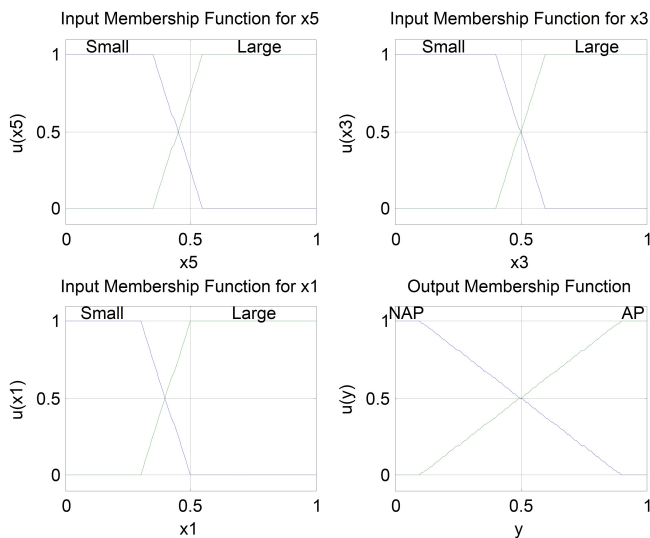


Figure 5. Input and output membership functions of fuzzy logic-based classifier from support vector machine. AP, anomalous propagation echo; NAP, non-anomalous propagation echo.

name the divided data as A, B, C, and D, respectively.

- Step 3: Implement SVM and ANN using the dataset A.
- Step 4: Apply dataset B to implemented SVM and ANN, described as B_1 and B_2 . The results can be considered as representations of each classifier.
- Step 5: Implement decision tree using B_1 and B_2 .
- Step 6: Verify performances of decision trees.
- Step 7: Derive a series of crisp rules from each decision

tree.

- Step 8: Convert crisp rules into fuzzy rules for constructing fuzzy logic-based classifiers.
- Step 9: Set input and output membership parameters and build fuzzy logic-based classifiers.

Figure 2 shows an example of decision tree for deriving a fuzzy inference system. We applied 5 properties as inputs of the SVM and ANN: a centroid altitude of the cluster (x_1), an average reflectivity data (x_2), a maximum reflectivity data (x_3), an average Doppler velocity (x_4), and a minimum Doppler velocity (x_5). In Section 3, we describe how the input variables are generated. And the decision tree indicates there is only 3 important inputs for separating anomalous propagation echo. Also, we can construct fuzzy rules using the decision tree as follows. The first fuzzy rules are derived from Figure 2, and the second fuzzy rules are derived from Figure 4.

- Rule 1: If x_1 is *small*, then y is *NOTAP*.
- Rule 2: If x_1 is *large* and x_5 is *small*, then y is *NOTAP*.
- Rule 3: If x_1 is *large* and x_5 is *large* and x_2 is *small*, then y is *AP*.
- Rule 4: If x_1 is *large* and x_5 is *large* and x_2 is *large*, then y is *NOTAP*.
- Rule 1: If x_5 is *large*, then y is *NOTAP*.
- Rule 2: If x_5 is *small* and x_3 is *small*, then y is *NOTAP*.
- Rule 3: If x_5 is *small* and x_3 is *large* and x_1 is *small*, then y is *NOTAP*.
- Rule 4: If x_5 is *small* and x_3 is *large* and x_1 is *large*, then y is *AP*.

From these rules, the SVM and the ANN consider different input as most significant variable according to the roots of the decision trees and the induced fuzzy rules. In case of the SVM, the centroid altitude of the cluster is most important variable. On the other hands, the minimum Doppler velocity is most important in case of the ANN. The common input variables are x_1 and x_5 , and the different input variables are x_2 and x_3 . Further, x_4 seems not significant influence because it is not shown in the trees and rules. Their input and output membership functions are generated as shown in Figures 4 and 5, respectively. The functions are trapezoidal shaped function.

3. Anomalous Propagation Echo

Due to properties of remote sensing device, the radar observation efficiency depends on the atmospheric condition. In other words, a beam path of weather radar can be changed by temperature, humidity, etc. The changed beam paths can be categorized as sub-refraction, normal refraction, super-refraction, and ducting according to its refractive index [16]. The sub-refraction phenomenon occurs when the radar beam is refracted toward opposite way of surface more than the normal refraction. It causes relatively low influence to weather forecasting. However, when the radar beam is deviated toward the surface by super-refraction or ducting, the resultant echo represents reflection of the ground or the sea surface which is not a meteorological target. It is called as an anomalous propagation echo. The weather radar computes altitude of observation targets consider as the normal refraction of the radar beam. Therefore, unexpected echoes could appear in the observation region of the weather radar by a surface scattering when the super-refraction or the ducting occurs.

It is one of the representative contamination source in the weather forecasting process because it induces a severe problem in quantitative precipitation estimation. The anomalous propagation echo should be removed from radar data because the echo originating from the surfaces can be misinterpreted as heavy precipitation in low altitude. In short, the refracted signals may lead to large overestimates of precipitation by the radar beam seeing surface instead of the atmosphere. Also, its location is difficult to predict. Furthermore, when the radar beam refracts toward the surface more severely, the intensity and extension of clutter areas can also change [17].

The entire proposed system is shown in Figure 6. The detailed sequence is described below. First of all, we need to clarify why we select corrected reflectivity (CZ) and Doppler velocity data (VR). There are several kinds of useful information in raw radar data such as spectrum width (SW) and uncorrected reflectivity (DZ). According to recent research for the anomalous propagation echo classification, the echo has following properties [18]: a near-zero radial velocity, a low spectrum width, a high texture of the reflectivity field, and so on. In this reason, we select the corrected reflectivity and a Doppler velocity data as input features.

Due to observation principles of the weather radar, the raw radar data follows the spherical coordinate. Therefore, in order to analyze the radar data intuitively, a coordinate conversion process should be applied from spherical to Cartesian. This

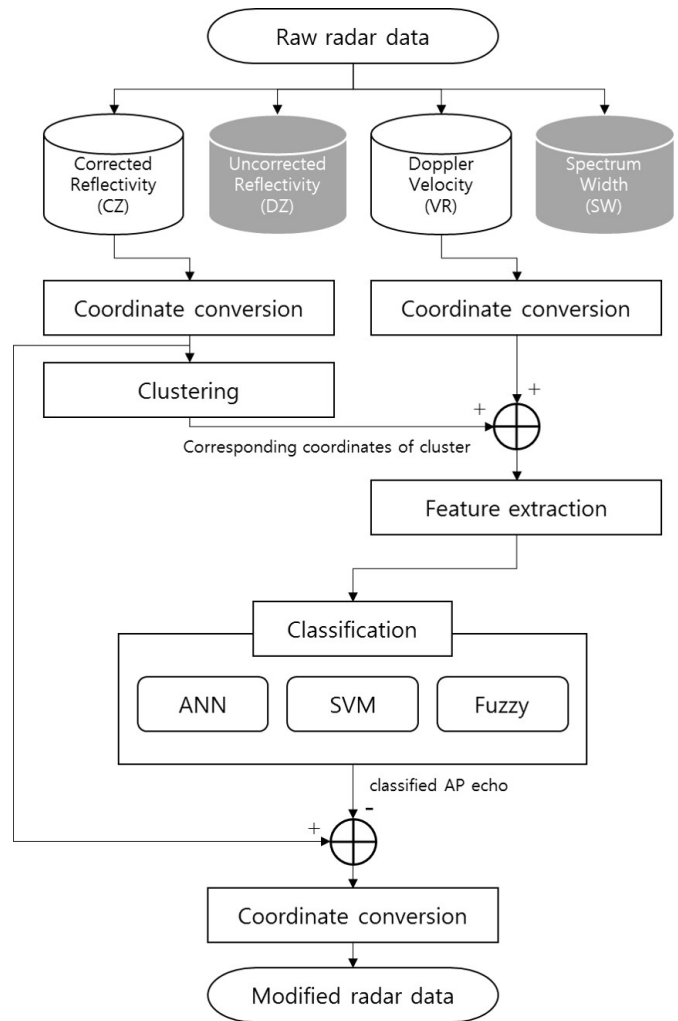


Figure 6. An actual implementation of decision tree. ANN, artificial neural network; SVM, support vector machine.

process makes to apply a clustering algorithm for grouping individual point data. Also, in order to find same locations, the coordinate conversion process is applied to both the corrected reflectivity and Doppler velocity data.

After the coordinate conversion, a spatial clustering method [19] which is one of hierarchical clustering methods is applied for grouping the reflectivity data. The clustered data is easier to deal with than the raw data because the radar data includes millions of data points. The proposed system uses statistical features derived from the clusters such as mean, minimum, maximum and its centroid position.

In feature extraction process, five properties are derived and used as inputs: centroid altitude of the cluster (x_1), average reflectivity data (x_2), maximum reflectivity data (x_3), average Doppler velocity (x_4), and minimum Doppler velocity (x_5). The reason why we select the centroid altitude of the cluster is that

the anomalous propagation echo appears in low altitude by its own properties.

After the feature extraction process, a classification method is applied. The detailed sequences are described in next section. Using the classifiers, each cluster is classified +1 or -1, which indicates that the selected cluster is anomalous propagation echo or others, respectively. The classified clusters which determined as an anomalous propagation echo are removed from the corrected reflectivity data.

After the removal process is done, the reverse coordinate conversion process is applied: from Cartesian to spherical coordinate. Finally, the processed radar data without the anomalous propagation echo is generated.

4. Experimental Results

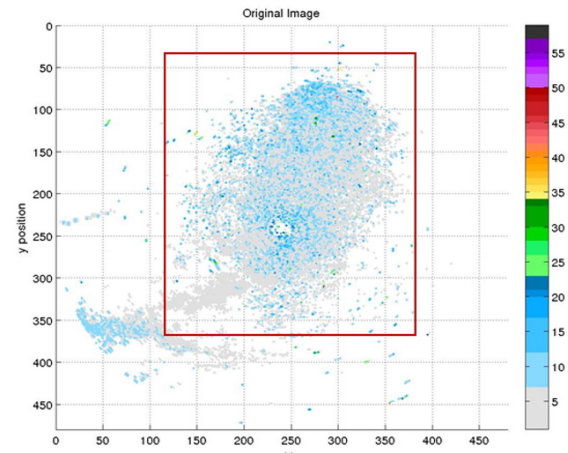
For the purpose of verifying the proposed system, we selected actual appearance cases in this paper for training and testing. Figures 7 and 8 show the classification results using the implemented fuzzy inference system with actual appearance cases of the anomalous propagation echo.

The case shown in Figure 7 indicates a case with small amount of precipitation echo in left lower side. As shown in Figure 7(a), a squared mark on the center region represents as the anomalous propagation echo. Figure 7(b) shows the radar image without the classified anomalous propagation echo by the proposed system. And Figure 7(c) describes the separated anomalous propagation echo.

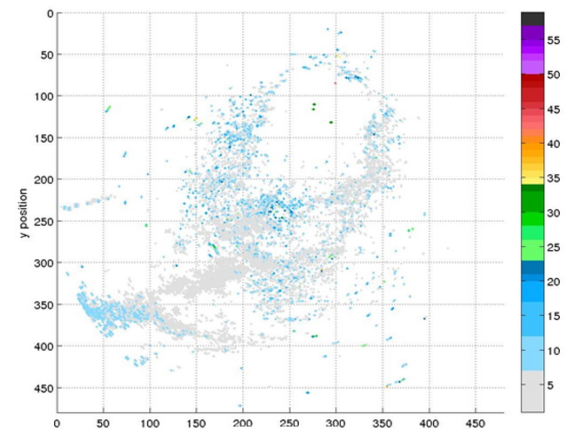
The case shown in Figure 8 indicates an independent appearance case of the anomalous propagation echo. As shown in Figure 8(a), the entire region in the squared marks represent as the anomalous propagation echo. Figure 8(b) shows the radar image without the classified anomalous propagation echo by the proposed method. And Figure 8(c) shows the anomalous propagation echo only.

From Figures 7 and 8, it is confirmed that the most of the regions of the anomalous propagation echo are removed in both cases: with and without precipitation echoes. In conclusion, the induced rule-based fuzzy inference system from the black-box models can be evaluated well according to these experiment results and accuracy comparison results.

There are several performance indexes for verifying classification methods such as accuracy, precision, and so on. In this paper, a confusion matrix is applied for calculating the accuracy as shown in equation (3).



(a)



(b)



(c)

Figure 7. An actual implementation of fuzzy logic-based classifier. (a) The anomalous propagation echo. (b) The radar image without the classified anomalous propagation echo by the proposed system. (c) The separated anomalous propagation echo.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

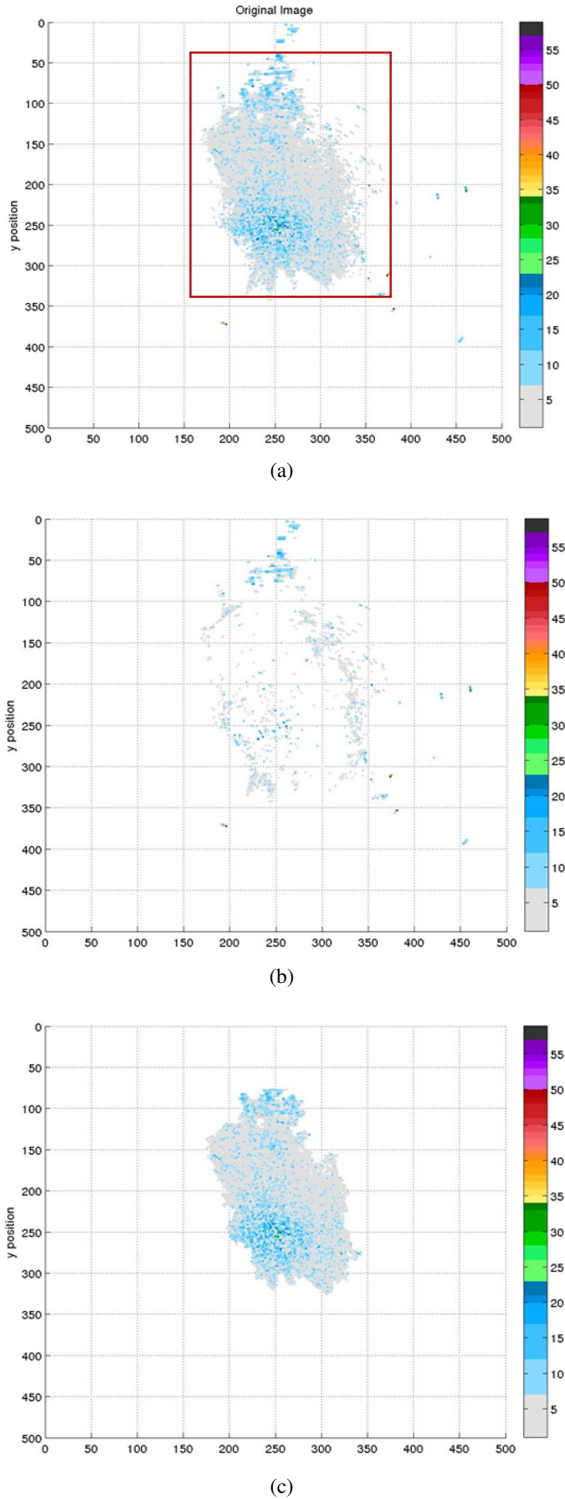


Figure 8. An actual implementation of fuzzy logic-based classifier. (a) The entire region of the anomalous propagation echo. (b) The radar image without the classified anomalous propagation echo by proposed method. (c) The anomalous propagation echo only.

Each parameter in equation (3) indicates as follow: TP for true positive, TN for true negative, FP for false positive, and FN

Table 1. Simulation and experimental system of the SVM parameters

	SVM	DT _{SVM}	Fuzzy _{SVM}
Site 1	94.39%	88.18%	91.51%
Site 2	91.27%	84.01%	87.52%
Site 3	89.11%	80.27%	83.43%

SVM, support vector machine; DT, decision tree.

Table 2. Simulation and experimental system of the ANN parameters

	ANN	DT _{ANN}	Fuzzy _{ANN}
Site 1	94.39%	88.69%	92.25%
Site 2	90.96%	85.42%	87.53%
Site 3	85.60%	81.95%	82.04%

ANN, artificial neural network; DT, decision tree.

for false negative. The true parameter indicates the anomalous propagation echo. And the false parameter is the other echoes. In this paper, we select three radar sites for experiments and evaluate accuracies.

The average accuracy of the SVM and the ANN classifier is shown in Tables 1 and 2, respectively. Due to the indirect approach using decision tree, the derived decision trees have slightly lower accuracy. But the induced fuzzy logic-based classifiers show better than the trees. The results indicate that flexible decision boundaries have beneficial effects on the accuracy, which the fuzzy logic-based classifier have.

5. Conclusions

In weather forecasting process, it is important to analyze the radar data accurately. Among the non-precipitation echoes, the anomalous propagation echo is one of the representative non-precipitation echo. This paper proposes the fuzzy inference system with induced rules from the SVM and the ANN. The five different properties derived by clustering algorithm are applied as inputs of the classifiers. We can conclude that the fuzzy inference system from the SVM can detect the anomalous propagation echo well.

Further proposed work is to improve accuracy of detecting the anomalous propagation echo. The induced membership function parameters should be optimized for improving accuracy. Also, the other classification method could be applied such as artificial neural network, naive Bayesian classifier, and

so forth. The empirical study is needed to select most appropriate algorithm for the anomalous propagation echo. Finally, the proposed system could be applied to other non-precipitation echoes.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2056958) and was supported by Global PhD Fellowship Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013-034596).

References

- [1] R. Andrews, J. Diederich, and A. B. Tickle, "Survey and critique of techniques for extracting rules from trained artificial neural networks," *Knowledge-Based Systems*, vol. 8, no. 6, pp. 373-389, 1995. [http://dx.doi.org/10.1016/0950-7051\(96\)81920-4](http://dx.doi.org/10.1016/0950-7051(96)81920-4)
- [2] T. Hill, L. Marquez, M. O'Connor, and W. Remus, "Artificial neural network models for forecasting and decision making," *International Journal of Forecasting*, vol. 10, no. 1, pp. 5-15, 1994. [http://dx.doi.org/10.1016/0169-2070\(94\)90045-0](http://dx.doi.org/10.1016/0169-2070(94)90045-0)
- [3] R. Fuller, *Neural Fuzzy Systems*. Turku, Finland: Abo Akademi University, 1995.
- [4] C. F. Lin and S. D. Wang, "Fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 464-471, 2002. <http://dx.doi.org/10.1109/72.991432>
- [5] A. J. Smola and B. Scholkopf, *Learning with Kernels*. Cologne, Germany: GMD-Forschungszentrum Informationstechnik, 1998.
- [6] N. Barakat and A. P. Bradley, "Rule extraction from support vector machines: a review," *Neurocomputing*, vol. 74, no.1-3, pp. 178-190, 2010. <http://dx.doi.org/10.1016/j.neucom.2010.02.016>
- [7] B. J. Taylor and M. A. Darrah, "Rule extraction as a formal method for the verification and validation of neural networks," in *Proceedings of IEEE International Joint Conference on Neural Networks (IJCNN'05)*, Montreal, Canadian, 2005, pp. 2915-2920. <http://dx.doi.org/10.1109/IJCNN.2005.1556388>
- [8] R. Setiono and H. Liu, "Symbolic representation of neural networks," *Computer*, vol. 29, no. 3, pp. 71-77, 1996. <http://dx.doi.org/10.1109/2.485895>
- [9] R. Setiono, "Extracting rules from neural networks by pruning and hidden-unit splitting," *Neural Computation*, vol. 9, no.1, pp. 205-225, 1997. <http://dx.doi.org/10.1162/neco.1997.9.1.205>
- [10] A. Gupta, S. Park, and S. M. Lam, "Generalized analytic rule extraction for feedforward neural networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, no. 6, pp. 985-991, 1999. <http://dx.doi.org/10.1109/69.824621>
- [11] T. A. Etchells and P. J. G. Lisboa, "Orthogonal search-based rule extraction (OSRE) for trained neural networks: a practical and efficient approach," *IEEE Transactions on Neural Networks*, vol. 17, no. 2, 2006. <http://dx.doi.org/10.1109/TNN.2005.863472>
- [12] N. H. Barakat and A. P. Bradley, "rule extraction from support vector machines: A sequential covering approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 6, pp. 729-741, 2007. <http://dx.doi.org/10.1109/TKDE.2007.190610>
- [13] X. Fu, C. J. Ong, S. Keerthi, G. G. Hung, and L. Goh, "Extracting the knowledge embedded in support vector machines," in *Proceedings of IEEE International Joint Conference on Neural Networks*, Budapest, Hungary, 2004. <http://dx.doi.org/10.1109/IJCNN.2004.1379916>
- [14] J. He, H. J. Hu, R. Harrison, P. C. Tai, and Y. Pan, "Rule generation for protein secondary structure prediction with support vector machines and decision tree," *IEEE Transactions on NanoBioscience*, vol. 5, no. 1, pp. 46-53, 2006. <http://dx.doi.org/10.1109/TNB.2005.864021>
- [15] N. Barakat and J. Diederich, "Learning-based rule-extraction from support vector machines: performance on

benchmark data sets,” in *Proceedings of the 14th International Conference on Computer Theory and applications (ICCTA2004)*, Alexandria, Egypt, pp. 1-8, 2004.

- [16] M. Steiner and J. A. Smith, “Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data,” *Journal of Atmospheric and Oceanic Technology*, vol. 19, no. 5, pp. 673-686, 2002. [http://dx.doi.org/10.1175/1520-0426\(2002\)019<0673:UOTDRS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0426(2002)019<0673:UOTDRS>2.0.CO;2)
- [17] M. A. Rico-Ramirez and I. D. Cluckie, “Classification of ground clutter and anomalous propagation using dual-polarization weather radar,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 1892-1904, 2008. <http://dx.doi.org/10.1109/TGRS.2008.916979>
- [18] C. Kessinger, S. Ellis, and J. Van Andel, “The radar echo classifier: a fuzzy logic algorithm for the WSR-88D,” in *Proceedings of the 3rd Conference on Artificial Intelligence Applications to the Environmental Science*, Long Beach, CA, pp. 1-11, 2003.
- [19] Y. H. Kim, S. Kim, H. Y. Han, B. H. Heo, and C. H. You, “Real-time detection and filtering of chaff clutter from single-polarization Doppler radar data,” *Journal of*

Atmospheric and Oceanic Technology, vol. 30, no. 5, pp. 873-895, 2013. <http://dx.doi.org/10.1175/JTECH-D-12-00158.1>



Hansoo Lee received the B.S. and M.S. degrees in Department of Electrical and Computer Engineering from Pusan National University, Busan, Korea, in 2010 and 2013, respectively, and is currently pursuing the Ph.D.

degree in Electrical and Computer engineering at Pusan National University, Busan, Korea. His present interests include intelligent system and data mining.

E-mail: hansoo@pusan.ac.kr



Sungshin Kim received his B.S. and M.S. degrees in Electrical Engineering from Yonsei University, Korea, in 1984 and 1986, respectively, and his Ph.D. degree in Electrical Engineering from the Georgia Institute

of Technology, USA, in 1996. He is currently a professor at the Department of Electrical and Computer Engineering, Pusan National University. His research interests include fuzzy logic controls, neuro fuzzy systems, neural networks, robotics, signal analysis, and intelligent systems.

E-mail: sskim@pusan.ac.kr