Research Article

Anomalous Propagation Echo Classification of Imbalanced Radar Data with Support Vector Machine

Hansoo Lee, Eun Kyeong Kim, and Sungshin Kim

Department of Electrical and Computer Engineering, Pusan National University, Busan 46241, Republic of Korea

Correspondence should be addressed to Sungshin Kim; sskim@pusan.ac.kr

Received 23 September 2015; Revised 30 November 2015; Accepted 10 January 2016

Academic Editor: Brian R. Nelson

Copyright © 2016 Hansoo Lee et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A number of technologically advanced devices, such as radars and satellites, are used in an actual weather forecasting process. Among these devices, the radar is essential equipment in this process because it has a wide observation area and fine resolution in both the time and the space domains. However, the radar can also observe unwanted nonweather phenomena. Anomalous propagation echo is one of the representative nonprecipitation echoes generated by an abnormal refraction phenomenon of a radar beam. Abnormal refraction occurs when the temperature and the humidity change dramatically. In such a case, the radar recognizes either the ground or the sea surface as an atmospheric object. This false observation decreases the accuracy of both quantitative precipitation estimation and weather forecasting. Therefore, a system that can automatically recognize an anomalous propagation echo from the radar data needs to be developed. In this paper, we propose a classification method for separating anomalous propagation echoes from the rest of the weather data by using a combination of a support vector machine classifier and the synthetic minority oversampling technique, to solve the problem of imbalanced data. By using actual cases of anomalous propagation we have confirmed that the proposed method provides good classification results.

1. Introduction

Weather forecasting provides important information that affects a variety of fields ranging from those of personal interest to those that provide economic benefits to companies. A ground-based weather radar system not only provides essential data but also has many other advantages: fine resolution in the space domain, information in near-real time, and a wide observation range in remote areas [1, 2]. Hence, we need to analyze radar data precisely for accurate weather forecasting.

Further, because a radar system cannot selectively observe meteorological phenomena such as snow, rain, and clouds, unwanted nonprecipitation echoes are inevitably included inside the observation data. Obviously, such observation data need to be examined by meteorologists because such observation data are difficult to directly apply to an analysis without verification. Moreover, the reliability of the forecasting result may change according to the skill level of meteorologists if the prediction system depends on the expert's decision. Therefore, we need further research

on the implementation of a system that can automatically distinguish nonprecipitation echoes from radar data

Certain representative types of nonprecipitation echoes are as follows: sea clutter [3, 4], sun strobe echoes and radial interference echoes [5], chaff echoes [6, 7], and anomalous propagation echoes [8–21]. Among them, an anomalous propagation echo is a radar signal generated by an abnormally refracted radar beam due to the temperature and humidity distribution in the atmosphere. When the refracted radar beam moves toward the ground or sea surface, the radar misinterprets some objects on the surface as meteorological objects. Because the size and the intensity of an anomalous propagation echo rely on the refractive index of the radar beam, it is difficult to accurately estimate its occurrence position and reflectivity.

There are several ongoing studies, across the world, on the separation of anomalous propagation echoes from radar data. Some of the representative research can be categorized as follows: case studies of anomalous propagation echoes in a specific radar observation area [9–11]; various classification algorithm application cases (Bayesian classifier [12, 13], neural

FIGURE 1: Radar data structure.

network [14–16], fuzzy logic classifier [17, 18], Markov chain [19], etc.); performance comparisons between classification algorithms [20]; and verification of classification algorithms with a number of occurrence cases [21].

In this paper, we propose a classification method that uses a support vector machine (SVM) to separate the anomalous propagation echoes. The SVM method derives the best hyperspace that divides the given data into two groups and is selected because of its following advantages: great performance in practical applications such as artificial neural network, rapid classification using relatively less data, problem compensations such as over fitting, and local optimization search [22–27].

However, when the class imbalance is not considered, minor class data will be ignored and major class data will be focused upon if a learning algorithm is applied. In general, one class of data exists more than the others in real-world applications. In particular, when a binary classification is performed, there may be a significant difference between major and minor classes. Several techniques, such as upsampling, downsampling, feature selection, threshold, and synthetic minority over sampling technique (SMOTE), have been proposed for solving this problem. Among them, the SMOTE method [28–30], which generates the synthetic data of the minority class in the feature space with the k -nearest neighbor algorithm (k -NN) methods, is applied in many practical fields and has shown successful results. Therefore, we select the SMOTE method for handling the class imbalance problem. As a result, the class imbalance problem is minimized by using the SMOTE method in this study.

Furthermore, the proposed classification method can be combined with various types of software. For example, there is a WRADLIB [31] software for analysis of radar data which provides a fuzzy echo classification and clutter identification. The proposed classification method can be used with the software and can create a synergy effect as follows: first, the fuzzy echo classification method can be replaced by the proposed SVM with the SMOTE method; second, the fuzzy

rules from the fuzzy echo classification method can be verified because it is possible to extract fuzzy rules from the proposed SVM with the SMOTE classifier; third, it is also possible to select most important inputs by using the verification and comparison of the performance. It allows adding or reducing the number of fuzzy rules. Thus, these mentioned facts can allow establishing more accurate and efficient system.

The remainder of this paper is organized as follows: in Section 2, the occurrence principles and properties of the anomalous propagation echo are elucidated. The SVM method for classification is described in Section 3. In Section 4, the SMOTE method is explained to overcome the imbalanced dataset problem. The verifications and experiments of the SVM classifier are described by using actual cases in Section 5. Finally, in Section 6, the conclusions and future works are discussed.

2. Anomalous Propagation

In this section, background knowledge is explained for understanding the properties of anomalous propagation echoes. This section is organized as follows: radar data structure, occurrence principles of anomalous propagation echoes, expert's rule for detection, learning data generation, and overview of the proposed system.

2.1. Radar Data Structure. Figure 1 shows the universal format (UF) file. UF file can be accessed systematically by using radar software library (RSL) supported by the tropical rainfall measuring mission (TRMM) [32].

Weather radars can observe and store various atmospheric phenomena because of the development of scientific technology. Various types of information can be obtained from radar data, such as reflectivity, Doppler velocity, and spectrum width. Therefore, only specific observation values are selected for an analysis. We chose the corrected reflectivity and the Doppler velocity for detecting an anomalous propagation echo in this study. Corrected reflectivity data

FIGURE 2: Principle of radar data observation.

FIGURE 3: Occurrence principle of anomalous propagation echo.

are preprocessed data unlike uncorrected reflectivity data in which ground echoes are removed from the raw data beforehand.

During the observation process of a radar system, the system maintains a constant elevation angle and varies the azimuth angle. When the azimuth angle reaches 360° , the elevation angle is changed. After several changes in the elevation angle, overlapping cone-shaped data of spherical coordinates are generated as shown in Figure 2. In order to analyze these data intuitively and apply the proposed method easily, a radar data sweep is performed to convert the related spherical coordinates into Cartesian coordinates by using RSL.

2.2. Occurrence Principles of Anomalous Propagation Echo. On the basis of the characteristics of a remote observation device using electromagnetic waves, we can change a propagation pathway depending on atmospheric conditions such as temperature, pressure, and humidity. The changed propagation pathway influences the observational efficiency and is classified according to the refractive index as follows: subrefraction, normal refraction, superrefraction, and ducting.

A subrefraction phenomenon refracts the propagation pathway in a direction opposite to that of the surface as compared to normal refraction. Because of this phenomenon, the radar underestimates the altitude of the observed object. In contrast, a superrefraction phenomenon refracts the propagation pathway in the surface direction as compared to normal refraction. Thus, the radar overestimates the altitude of the observed object. In particular, a ducting phenomenon traps the propagation pathway between a specific atmospheric layer and the surface by severely refracting the electromagnetic waves of the radar.

Figure 3 illustrates the different propagation pathways. An echo with intense reflectivity appears upon the surface scattering of radar propagation waves when superrefraction

or the ducting phenomenon occurs; this is generally referred to as an anomalous propagation echo. The anomalous propagation echo should be removed for accurate forecasting because it causes errors in the quantitative precipitation estimation process.

Some actual cases of anomalous propagation echoes are shown in Figure 4 in Korea. Figure 4(a) illustrates the case of an anomalous propagation echo that was generated along with a precipitation echo and a sun strobe echo on November 2, 2011. The anomalous propagation echo is indicated by a curved arrow. Figure 4(b) illustrates the case of an anomalous propagation echo generated along with a second echo, denoted as “2nd” on March 28, 2012. Figure 4(c) shows the case of an anomalous propagation echo that was generated along with a stratiform echo, denoted as “Up” on April 14, 2012.

Several problems may occur without the decision-making by meteorologists, such as incorrect estimation of distribution of precipitation echoes or quantitative precipitation in Figure 4(a) and miscalculation of precipitation echoes in Figures 4(b) and 4(c). False weather forecasting on the basis of an anomalous propagation echo could lead to inconveniences or damages to individuals as well as the entire country either directly or indirectly. Therefore, in order to prevent inaccurate weather forecasting, anomalous propagation echoes should be separated from the rest of the radar data and removed.

2.3. Expert's Rule for Detection. There are several intricate rules for removing anomalous propagation echoes. Meteorologists select from these intricate rules considering many complex requirements. The most important rule is that meteorologists or automated systems should not misclassify a precipitation echo as a nonprecipitation echo. The rules that apply representatively in an actual anomalous propagation echo occurrence case can be summarized as follows.

Rule 1. The location change should be small when the echo is moved (Doppler velocity: 0 m/s on ground, and more than 0 m/s above sea).

Rule 2. Appearance characteristics should be slightly different in the on-ground and on-sea surface cases.

Rule 3. Because of the low altitude, it is difficult to separate low altitude rain and snow echoes.

Rule 4. The reflectivity distribution is discontinuous in the vertical and horizontal directions.

According to these rules of the experts shown above, Doppler velocity and reflectivity are essential data. Therefore, in this study, we statistically utilize the Doppler velocity and reflectivity data relevant to Rules 1 and 4 as inputs of an SVM classifier.

A large number of data points have to be considered when we carry out an analysis using Cartesian coordinates, as shown in

$$\text{Number of Data Points} = (\text{radius})^2 \times (\text{altitude}). \quad (1)$$

FIGURE 4: Anomalous propagation echo: (a) November 2, 2011, (b) March 28, 2012, and (c) April 14, 2012.

For example, when the radar data have a 240 km observation radius with a 1 km resolution, and a 10 km altitude with a 250 m resolution, theoretically, 9,446,440 data points should be considered for the analysis. Although such a high number of data points has not been noted in actual observation cases, selecting each data point as an individual input of the classifier is an inefficient process because a large-sized echo can have more than 1 million data points.

2.4. Learning Data Generation. In this study, the radar data is obtained by a single polarization radar. By its operational properties, there are relatively limited amount of input variables compared to a dual polarization radar, especially related to vertically polarized wave. Although it is possible to use vertical structure of the observed reflectivity in the single polarization radar data, it has the undesired effect of eliminating echoes from shallow stratiform precipitation [13]. Also, according to a previous research [17], clutter echoes are constant in neither intensity, vertical extent, nor location because of the variation of atmospheric conditions. Further, there is a research for improving weather radar precipitation estimates by combining two types of radars [33]. Considering that the single polarization radars are set up throughout the world and are used in actual meteorologically related fields, it is essential to remove the anomalous propagation echo because it disturbs the precipitation estimation process significantly as mentioned before. Therefore, we apply a clustering method to replace some properties which can be used as input variables such as vertical structure of the observed reflectivity, and so forth.

A learning data generation process is performed by a spatial clustering algorithm [6]. It converts a given raw data into a labeled data. The labeled data can provide some statistical properties shown in Table 1. The meanings of the properties are as follows: cluster identification number, which is automatically generated identification number for separating out different clusters; maximum and mean corrected reflectivity by aforementioned Rule 4; maximum, minimum, and mean Doppler velocity by aforementioned Rule 1. In this

study, we define $\mathbf{x}_c$ vector in a feature space in order to apply the properties to the proposed classification method, as shown in

$$\mathbf{x}_c = \{x_{c1}, x_{c2}, x_{c3}, x_{c4}, x_{c5}\},$$

$$x_{c1} = \frac{1}{N_c} \sum_{n_c=1}^{N_c} CZ_c(n_c),$$

$$x_{c2} = \max_{n_c \in N_c} CZ_c(n_c),$$

$$x_{c3} = \frac{1}{N_c} \sum_{n_c=1}^N VR_c(n_c),$$

$$x_{c4} = \min_{n_c \in N_c} VR_c(n_c),$$

$$x_{c5} = \max_{n_c \in N_c} VR_c(n_c),$$

$$\{c \mid c \in C, C : \text{identification number of clusters}\},$$

$$\{n_c \mid n_c \in N_c, N_c :$$

$$\text{number of elements in specific cluster } c\},$$

where c which is one of elements in C indicates an identification number of cluster, n_c is number of elements in specific cluster c , CZ_c is elements of corrected reflectivity in specific cluster c , and VR_c indicates elements of Doppler velocity element in specific cluster c . Elements of the defined vector $\mathbf{x}_c$ are mean and maximum corrected reflectivity of the cluster c , mean, minimum, and maximum Doppler velocity of the cluster c , sequentially. Further, the minimum corrected reflectivity is not taken into account as an input of the SVM classifier because the minimum value of the anomalous propagation echo is inconsistent in actual cases.

As a last process, we need to label the cluster whether it is an anomalous propagation echo or not. We have a series of raw radar data with expert's decision in figure form, which contains the region where an anomalous propagation echo

TABLE 1: Learning dataset from derived clusters.

Cluster ID	Mean reflectivity	Maximum reflectivity	Minimum Doppler velocity	Mean Doppler velocity	Maximum Doppler velocity	Class
0	6.26	53.00	0.00	74.30	255.00	AP
119	10.30	39.00	24.00	239.21	255.00	AP
186	16.61	30.00	1.00	5.08	253.00	NOTAP
465	8.39	19.00	0.00	26.55	254.00	AP
473	13.24	29.00	3.00	8.24	226.00	NOTAP
686	7.41	17.00	0.00	0.25	252.00	NOTAP

FIGURE 5: Overview of the proposed system.

exists or not. Therefore, after visualizing clusters separately and comparing between the visualized region and the figure containing expert's decision, we decide a class of the clusters.

2.5. Overview of Proposed System. The proposed anomalous propagation echo recognition and removal system is shown in Figure 5. This system consists of one input (raw radar data), five main steps (coordinate conversion, clustering, data balancing, classification, and coordinate restoration), and one output (modified radar data with the anomalous propagation echo removed).

First, the coordinate conversion process is performed because the raw radar data follow spherical coordinates on the basis of the observation principle; these coordinates need to be converted into Cartesian coordinates for an intuitive analysis and application of the proposed algorithm. Secondly, the spatial clustering algorithm, which is a type

of hierarchical clustering, is applied in order to categorize the radar data and extract statistical properties. Thirdly, the SMOTE method is applied for minimizing the influences of the data imbalance problem, which occurs frequently in real-world applications. The SMOTE method will be elucidated in Section 3. Fourthly, the classification process using the SVM algorithm is performed to identify the anomalous propagation echo by utilizing the input data generated using the clustering and data balancing methods. The SVM algorithm will be discussed in Section 4. Finally, in order to maintain the consistency of the file format in the data storage, the coordinate restoration process is performed; this process converts Cartesian coordinates into spherical coordinates.

3. Synthetic Minority Oversampling Technique

It is essential to secure a sufficient amount of learning data in order to identify an anomalous propagation echo accurately in weather radar data by utilizing supervised learning. However, a classified implemented in a real-world application shows low performance because most of the learning methods assume that the dataset has a balanced distribution according to its classes. In particular, the problem occurs when the dataset has a high-dimensional feature space.

There are several methods to overcome these problems such as oversampling, undersampling, threshold, and SMOTE. Among them, SMOTE is an advanced version of sampling with replacement method. In other words, it is a technique for balancing between major and minor classes by generating synthetic minority data using the k -NN algorithm in the feature space.

As shown in (3), synthetic data are generated by using k -NN and a random process as follows: first, k number of data items are selected by the k -NN method from the minority class data. Next, two of the selected data items are chosen randomly. Finally, synthetic data are created between the two chosen data items by multiplying them with random values. In the following equation, C represents the coordinates and P indicates the probability distribution:

$$C_{\text{synthetic}} = C_{\text{origin}} + (C_{\text{origin}} - C_{\text{nth neighbor}}) P_{\text{uniform}}(x). \quad (3)$$

There are several representative and frequently chosen probability distributions such as Gaussian, Poisson, and uniform distributions. In this study, a uniformly distributed random

FIGURE 6: SMOTE algorithm.

variable is used, and its probability density function is described in

$$P_{\text{uniform}}(x) = \begin{cases} 1, & \text{for } x \in [0, 1], \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Figure 6 shows the principle of the SMOTE method. The triangular shape indicates the majority class data, and the rectangular shape represents the minority class data. We have confirmed that two novel types of synthetic data are generated, denoted by the filled rectangular shape.

For the purpose of verifying the usefulness of the SMOTE method, we use actual radar data. Figure 7 shows the data distribution between reflectivity and Doppler velocity: The asterisk denotes an anomalous propagation echo (AP), dot a nonanomalous propagation echo (NAP), and the cross synthetic AP data. As shown in Figure 7(a), outliers and overlapped data distributions can make the supervised classification method inaccurate.

The number of AP data items is 190, and the number of NAP data items is 261 in the case illustrated in Figure 7(a); thus, the ratio between AP and NAP is about 1:1.3. Further, the total number of AP data items is 501, and that of NAP data items is 1110 in three radar sites; in this case, the ratio between AP and NAP is 1:2.2. Thus, in order to obtain a more accurate SVM classifier, the dataset needs to be balanced for learning. As shown in Figure 7(b), the SMOTE algorithm creates synthetic data. Consequently, it is confirmed that the influence of outliers can be minimized, and a supervised classification method based on the balanced dataset can be implemented using the SMOTE algorithm.

4. Support Vector Machine

An SVM method 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, prediction, and estimation processes by using a relatively low amount of the given learning data [22–27].

As shown in Figure 8, the SVM method can find a separable hyperplane with the maximum margin from data having two classes. Assume a set S of labeled training data for learning the SVM method $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)$, where l denotes the number of data items; further, the set contains a predefined vector $\mathbf{x}$ and its respective class, where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \{-1, +1\}$ for $i = 1, \dots, l$. Here, $y = +1$ represents the anomalous propagation echo and $y = -1$ indicates the other echoes. As in most cases, there are many restrictions for finding a suitable hyperplane in its own input space. Thus, we need to find a suitable hyperplane in a higher-dimensional feature space than the input space in order to solve the abovementioned problem.

Let $\mathbf{z} = \varphi(\mathbf{x})$ denote the corresponding feature space vector. Here, φ denotes a function that can map the input space $\mathbb{R}^n$ to a higher-dimensional feature space $\mathbb{Z}$. Next, we aim to find a hyperplane that satisfies

$$\mathbf{w} \cdot \mathbf{z} + b = 0. \quad (5)$$

Equation (6) indicates the linear discriminant function form for $\mathbf{x}$, where $\mathbf{w} \in \mathbb{Z}$ and $b \in \mathbb{R}$:

$$f(\mathbf{x}_i) = \text{sign}(\mathbf{w} \cdot \mathbf{z}_i + b) = \begin{cases} 1, & \text{if } y_i = +1, \\ -1, & \text{if } y_i = -1. \end{cases} \quad (6)$$

If S is linearly separable, $(\mathbf{w}, b)$ can be described as follows:

$$\begin{aligned} (\mathbf{w} \cdot \mathbf{z}_i + b) &\geq 1, & \text{if } y_i = +1, \\ (\mathbf{w} \cdot \mathbf{z}_i + b) &\leq -1, & \text{if } y_i = -1, \\ i &= 1, \dots, l. \end{aligned} \quad (7)$$

Further, a unique optimal hyperplane that has the maximized margin can be derived by using two different classes of training data from the linearly separable set S . However, S is not linearly separable in most of the real-world applications. Considering that the radar is also not linearly separable, we can rewrite (7) as (8), which allows a classification violation:

$$y_i (\mathbf{w} \cdot \mathbf{z}_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l. \quad (8)$$

The nonnegative and nonzero ξ_i in (8) denote the fact that $\mathbf{x}_i$ does not satisfy (7). Consequently, $\sum_{i=1}^l \xi_i$ can be analyzed as a misclassification result. To find the optimal hyperplane in the linearly inseparable case, (8) can be changed to

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^l \xi_i \\ &\text{subject to} \quad y_i (\mathbf{w} \cdot \mathbf{z}_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l \\ &\quad \quad \quad \xi_i \geq 0, \quad i = 1, \dots, l. \end{aligned} \quad (9)$$

In (9), C denotes a constant that can be considered to be a regularization parameter because C is the only free

FIGURE 7: Radar data distribution: (a) without SMOTE algorithm and (b) with SMOTE algorithm.

FIGURE 8: SVM algorithm.

parameter in the SVM formulation. The balance between the classification violation and the margin maximization is adjusted by regulating C [25].

The problem in (9) is a quadratic programming problem. It can be solved by constructing a Lagrangian and transforming it into a dual, where $\alpha = (\alpha_1, \dots, \alpha_l)$ is the vector of the nonnegative Lagrange multiplier associated with the constraint in

$$\begin{aligned} \text{minimize } W(\alpha) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \mathbf{z}_i \cdot \mathbf{z}_j \\ \text{subject to } \sum_{i=1}^l \alpha_i y_i &= 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l. \end{aligned} \quad (10)$$

To construct the optimal hyperplane $\bar{\mathbf{w}} \cdot \mathbf{z} + \bar{b} = 0$, $\bar{b}$ can be determined from the Kuhn-Tucker conditions and $\bar{\mathbf{w}}$ can be derived by using

$$\bar{\mathbf{w}} = \sum_{i=1}^l \bar{\alpha}_i y_i \mathbf{z}_i. \quad (11)$$

Now, the decision function is generalized from (6) and (11) as shown in

$$f(\mathbf{x}_c) = \text{sign}(\mathbf{w} \cdot \mathbf{z} + b) = \text{sign}\left(\sum_{i=1}^l \bar{\alpha}_i y_i \mathbf{z}_i \cdot \mathbf{z} + b\right). \quad (12)$$

As there is no prior knowledge about φ , it appears that (10) and (12) cannot be computed. However, prior knowledge is not required by the SVM. As shown in (13), only the kernel function $K(\cdot, \cdot)$ is required for computing the dot product between the data points in feature space $\mathbb{Z}$:

$$\mathbf{z}_i \cdot \mathbf{z}_j = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j). \quad (13)$$

If the function satisfies Mercer's theorem, its value can be used as the dot product. Thus, the function that satisfies Mercer's theorem can be used as the kernel function. To implement the SVM classifier in this study, a radial basis function kernel is used as shown in

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right). \quad (14)$$

FIGURE 9: Overview of proposed SVM algorithm.

In addition to the linear separating hyperplane case shown in (10) and (12), a nonlinear separating hyperplane is shown in

$$\begin{aligned}
 &\text{minimize } W(\alpha) \\
 &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\
 &\text{subject to } \sum_{i=1}^l \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l
 \end{aligned} \quad (15)$$

and its decision function is shown in

$$\begin{aligned}
 f(\mathbf{x}) &= \text{sign}(\mathbf{w} \cdot \mathbf{z} + b) \\
 &= \text{sign}\left(\sum_{i=1}^l \bar{\alpha}_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right).
 \end{aligned} \quad (16)$$

In addition to the linear separating hyperplane case shown in (10) and (12), a nonlinear separating hyperplane is shown in (15) and its decision function is shown in (16).

In this study, we illustrate the process of obtaining modified data from raw data in Figure 9 with derived features shown in (2). The elements of (2) can be represented as

$$\begin{aligned}
 \mathbf{x}_c &= \{x_{c1}, x_{c2}, x_{c3}, x_{c4}, x_{c5}\} \in \mathbb{R}^5 \\
 x_{c1}: &\text{mean corrected reflectivity of cluster } c \\
 x_{c2}: &\text{maximum corrected reflectivity of cluster } c \\
 x_{c3}: &\text{minimum Doppler velocity of cluster } c \\
 x_{c4}: &\text{mean Doppler velocity of cluster } c \\
 x_{c5}: &\text{maximum Doppler velocity of cluster } c.
 \end{aligned} \quad (17)$$

By using the defined vector $\mathbf{x}_c$ in (17) and radial basis kernel function in (14), we can establish the decision function of the given cluster as shown in

$$\begin{aligned}
 f(\mathbf{x}_c) &= \text{sign}(\mathbf{w}_c \cdot \mathbf{z}_c + b_c) \\
 &= \text{sign}\left(\sum_{i=1}^l \bar{\alpha}_i y_{ci} \exp\left(-\frac{\|\mathbf{x}_{ci} - \mathbf{x}_c\|^2}{2\sigma^2}\right) + b_c\right).
 \end{aligned} \quad (18)$$

When the vector $\mathbf{x}_c$ in (17) is applied to the decision function in (18), the result is generated as +1 or -1 indicating the anomalous propagation echo or not, sequentially.

5. Experimental Results

Real anomalous propagation echo detection cases were used as data to verify the performance of the proposed SVM classifier. The actual radar data have a class-imbalanced distribution, as discussed in Section 3, and the learning data are constructed by using the SMOTE method.

The target location of the radar is the central region of Korea, specifically Baekryeongdo, Mt. Gwangdeok, and Gangneung. There are several radar sites in Korea, but we have selected these three specific radar sites for the following reasons: first, the Baekryeongdo radar is located on the Baekryeong Island, which does not have any high mountains around it. Therefore, the data from the Baekryeongdo radar are chosen because of the representative characteristics of anomalous propagation echoes on the sea surface. Second, the Gwangdeoksans radar is located on Mt. Gwangdeok; this location is surrounded by high mountains. Hence, the data from the Gwangdeoksans radar are chosen because of the representative characteristics of anomalous propagation echoes on the ground surface. Finally, the Gangneung radar is located in a relatively complex position: the left side of its observation area is blocked by high mountains, and

FIGURE 10: Experimental results: Baekryeongdo radar (a) image of original data, (b) image without anomalous propagation echoes, and (c) image of anomalous propagation echoes; Gwangdeoksan radar (d) image of original data, (e) image without anomalous propagation echoes, and (f) image of anomalous propagation echoes; Gangneung radar (g) image of original data, (h) image without anomalous propagation echoes, and (i) image of anomalous propagation echoes.

the right side opens to the ocean. Therefore, the data from the Gangneung radar are chosen because of the specific characteristics of anomalous propagation echoes that can be observed on both the sea surface and the earth surface.

Figure 10 illustrates the results of the SVM classifier utilizing learning data balanced by the SMOTE method. Figure 10(a) shows an anomalous propagation echo occurrence case from the Baekryeongdo radar. Precipitation echoes are not shown, and the anomalous propagation echo appears in a square area on the left side. Figure 10(d) shows an anomalous propagation echo occurrence case from the Gwangdeoksan radar. There are precipitation echoes on the top left.

However, the other echoes in the square area are radially formed from the anomalous propagation echo. Figure 10(g) shows an anomalous propagation echo occurrence case from the Gangneung radar. Precipitation echoes are not shown, and the anomalous propagation echo appears in the square area from the mountains to the coast.

In all cases, the occurrence of the anomalous propagation echo is determined by a meteorologist. Figures 10(b), 10(e), and 10(h) show radar images without an anomalous propagation echo, obtained by using the SVM classifier. Figures 10(c), 10(f), and 10(i) show the radar images classified as images containing anomalous propagation echoes by using

TABLE 2: Confusion matrix.

		Predicted	
		Positive	Negative
Actual	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

only the SVM classifier. On the basis of Figure 10, we confirm that the proposed SVM classifier with the SMOTE method successfully separates the anomalous propagation echoes for different occurrence cases.

In order to compare the results objectively, the performance of the SVM classifiers (one used along with the SMOTE method and the other, without) is evaluated by using a confusion matrix [34]. Also, we select an artificial neural network algorithm (ANN) for comparison, which is also one of the most popular supervised classifiers with remarkable classification performance. A confusion matrix as shown in Table 2 contains information about the actual and the predicted classifications done by a classification system.

By using a confusion matrix, we can calculate the classification accuracy as shown in

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \times 100 (\%). \quad (19)$$

Further, sensitivity and specificity are considered in this study for the purpose of performance evaluation including accuracy. According to meteorologists, as mentioned earlier, the most important rule is that meteorologists or algorithms should not misclassify a precipitation echo as a nonprecipitation echo. In other words, both the sensitivity described in (20) and the specificity described in (21) need to be used as the performance evaluation factors for actual field work:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 (\%), \quad (20)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \times 100 (\%). \quad (21)$$

Table 3 shows the experimental results of the k -fold cross-validation method used for an accurate performance comparison. Considering the parameters of data size and efficiency, we set k to 5. According to Table 3, the SVM classifier used along with the SMOTE method has an average accuracy of 85.02% and the SVM classifier used without the SMOTE method has an average accuracy of 90.01%. In other words, the experimental results reveal that the SVM classifier used along with the SMOTE method can obtain a higher average accuracy than that used without the SMOTE method by about 4.99%.

Additionally, an average accuracy of the ANN classifier used along with the SMOTE method is 82.23% and the ANN classifier used without the SMOTE method has an average accuracy of 85.51%. These results indicate that the SMOTE method can slightly increase the accuracy of the classifier.

Also, the SVM classifier used along with the SMOTE method has an average sensitivity of 85.78%, which is slightly lower, by about 1.66%, than that of the SVM classifier used

without the SMOTE method. But the sensitivity of the SVM classifier used along with the SMOTE method has a better average sensitivity about 9.52% compared to the ANN classifier and 0.41% compared to the ANN classifier with the SMOTE method.

Furthermore, the average specificity of the SVM classifier used along with the SMOTE method is 91.96%, which is higher than that of the SVM classifier used without the SMOTE method by about 9.22%. In addition, the average specificity of the SVM classifier used along with the SMOTE method is higher than the ANN classifier with and without the SMOTE method by about 16.01% and 13.28%, respectively.

In short, the proposed SVM classifier with the SMOTE method shows better accuracy and specificity and similar level of sensitivity compared to the SVM classifier without the SMOTE method. Further, the proposed classifier shows better performance in all aspects than the ANN classifier with and without the SMOTE method.

6. Conclusion

In this paper, we propose a classification method for separating the anomalous propagation echoes that are generated by an abnormally refracted radar beam because of the temperature and humidity distribution in the atmosphere, from the rest of the weather radar data. Because the size and the intensity of an anomalous propagation echo depend on the refractive index of the radar beam, it is difficult to estimate its occurrence position and reflectivity. Further, such an echo adversely affects the quantitative precipitation estimation and decreases the weather forecasting accuracy. In order to obtain an accurate forecasting result, the echo needs to be recognized and removed from radar data. Thus, in this paper, we propose a classification method for separating the anomalous propagation echo from the rest of the weather radar data by using an SVM classifier. Further, in this study, we use the SMOTE method to solve the problem of an imbalanced dataset, which occurs in a case study of the anomalous propagation echo, and enhance classification performance. The combination of the SVM classifier and the SMOTE method increases specificity, reduces false alarms, and improves quality control compared to the standalone SVM classifier and ANN classifiers which are both combined and standalone. For a rational selection of anomalous propagation echo cases, three radar systems located in different places, namely, one on an island, another surrounded by high mountains, and the other located between mountains and the ocean, are chosen. As a result, we conclude that the proposed method shows good performance with respect to recognizing and removing an anomalous propagation echo.

As an extension of the present work, future research will focus on the optimization of the classifier on the basis of the observation characteristics for different radar locations. Because the sea surface moves continuously, properties of an anomalous propagation echo are different on the sea surface and on the ground surface. Therefore, it would be inefficient to apply the same classifier in different environments. The abovementioned future work would solve this problem and

TABLE 3: Comparison of experimental results.

Radar site	BRI				GDK				GNG			
	SVM	SVM + SMOTE	ANN	ANN + SMOTE	SVM	SVM + SMOTE	ANN	ANN + SMOTE	SVM	SVM + SMOTE	ANN	ANN + SMOTE
Classification methods												
Accuracy	92.02%	95.83%	92.91%	95.27%	81.32%	86.38%	84.16%	84.37%	81.72%	87.81%	69.62%	76.89%
Sensitivity	91.38%	94.25%	74.83%	89.35%	83.93%	81.10%	82.13%	86.04%	87.01%	81.99%	71.81%	80.73%
Specificity	94.87%	98.89%	93.17%	94.06%	78.64%	91.88%	80.65%	81.44%	74.70%	94.02%	54.02%	60.55%

enhance the classifier performance. Further, we plan to apply the proposed method to the recognition of other anomalous propagation echoes, such as sea clutter and chaff echoes.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

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 Ph.D. Fellowship Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013-034596).

References

- [1] R. E. Rinehart, *Radar for Meteorologists*, University of North Dakota, Office of the President, 1991.
- [2] Met Office, National Meteorological Library and Archive Fact Sheet No. 15—Weather Radar, 2015, <http://www.metoffice.gov.uk/learning/library/publications/factsheets>.
- [3] L. Rosenberg, "Sea-spike detection in high grazing angle X-band sea-clutter," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 8, pp. 4556–4562, 2013.
- [4] S. Haykin, R. Barker, and B. W. Currie, "Uncovering nonlinear dynamics—the case study of sea clutter," *Proceedings of the IEEE*, vol. 90, no. 5, pp. 860–881, 2002.
- [5] V. Lakshmanan, A. Fritz, T. Smith, K. Hondl, and G. Stumpf, "An automated technique to quality control radar reflectivity data," *Journal of Applied Meteorology and Climatology*, vol. 46, no. 3, pp. 288–305, 2007.
- [6] 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.
- [7] X. Shao, H. Du, and J. Xue, "A new method of ship and chaff polarization recognition under rain and snow cluster," in *Proceedings of the IEEE International Workshop on Anti-counterfeiting, Security, Identification (ASID '07)*, pp. 142–147, IEEE, Fujian, China, April 2007.
- [8] S. Moszkowicz, G. J. Ciach, and W. F. Krajewski, "Statistical detection of anomalous propagation in radar reflectivity patterns," *Journal of Atmospheric and Oceanic Technology*, vol. 11, no. 4, pp. 1026–1034, 1994.
- [9] F. Mesnard and H. Sauvageot, "Climatology of anomalous propagation radar echoes in a coastal area," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 11, pp. 2285–2300, 2010.
- [10] M. Neuper and J. Handwerker, *Anomalous Propagation: Examination of Ducting Conditions and Anaprop Events in SW-Germany*, Seminar Work, Karlsruhe Institute of Technology, 2010.
- [11] J. Bech, A. Sairouni, B. Codina, J. Lorente, and D. Bebbington, "Weather radar anaprop conditions at a Mediterranean coastal site," *Physics and Chemistry of the Earth Part B: Hydrology, Oceans and Atmosphere*, vol. 25, no. 10–12, pp. 829–832, 2000.
- [12] J. A. Pamment and B. J. Conway, "Objective identification of echoes due to anomalous propagation in weather radar data," *Journal of Atmospheric and Oceanic Technology*, vol. 15, no. 1, pp. 98–113, 1998.
- [13] J. R. Peter, A. Seed, and P. J. Steinle, "Application of a bayesian classifier of anomalous propagation to single-polarization radar reflectivity data," *Journal of Atmospheric and Oceanic Technology*, vol. 30, no. 9, pp. 1985–2005, 2013.
- [14] R. B. Da Silveira and A. R. Holt, "An automatic identification of clutter and anomalous propagation in polarization-diversity weather radar data using neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 8, pp. 1777–1788, 2001.
- [15] M. Grecu and W. F. Krajewski, "An efficient methodology for detection of anomalous propagation echoes in radar reflectivity data using neural networks," *Journal of Atmospheric and Oceanic Technology*, vol. 17, no. 2, pp. 121–129, 2000.
- [16] M. Grecu and W. F. Krajewski, "Detection of anomalous propagation echoes in weather radar data using neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 1, pp. 287–296, 1999.
- [17] M. Berenguer, D. Sempere-Torres, C. Corral, and R. Sánchez-Diezma, "A fuzzy logic technique for identifying nonprecipitating echoes in radar scans," *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 9, pp. 1157–1180, 2006.
- [18] Y.-H. Cho, G. Lee, K.-E. Kim, and I. Zawadzki, "Identification and removal of ground echoes and anomalous propagation using the characteristics of radar echoes," *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 9, pp. 1206–1222, 2006.
- [19] B. Haddad, A. Adane, F. Mesnard, and H. Sauvageot, "Modeling anomalous radar propagation using first-order two-state Markov chains," *Atmospheric Research*, vol. 52, no. 4, pp. 283–292, 1999.
- [20] 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.
- [21] W. F. Krajewski and B. Vignal, "Evaluation of anomalous propagation echo detection in WSR-88D data: a large sample case study," *Journal of Atmospheric and Oceanic Technology*, vol. 18, no. 5, pp. 807–814, 2001.
- [22] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, Mass, USA, 2001.
- [23] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000.
- [24] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [25] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 464–471, 2002.
- [26] H. Lee and E. Kim, "Genetic outlier detection for a robust support vector machine," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 15, no. 2, pp. 96–101, 2015.

- [27] S.-Y. Lee, D. Ahn, M. Song, and K. Lee, "The classification of electrocardiograph arrhythmia patterns using fuzzy support vector machines," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 11, no. 3, pp. 204–210, 2011.
- [28] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [29] P. Jeatrakul, K. W. Wong, and C. C. Fung, "Classification of imbalanced data by combining the complementary neural network and SMOTE algorithm," in *Neural Information Processing. Models and Applications: 17th International Conference, ICONIP 2010, Sydney, Australia, November 22–25, 2010, Proceedings, Part II*, vol. 6444, pp. 152–159, Springer, Berlin, Germany, 2010.
- [30] J. Wang, M. Xu, H. Wang, and J. Zhang, "Classification of imbalanced data by using the SMOTE algorithm and locally linear embedding," in *Proceedings of the 8th International Conference on Signal Processing (ICSP '06)*, vol. 3, IEEE, Beijing, China, November 2006.
- [31] M. Heistermann, S. Jacobi, and T. Pfaff, "Technical Note: an open source library for processing weather radar data (wradlib)," *Hydrology and Earth System Sciences*, vol. 17, no. 2, pp. 863–871, 2013.
- [32] Tropical Rainfall Measuring Mission, NASA, September 2015, http://trmm-fc.gsfc.nasa.gov/trmm_gv/software/rsl/.
- [33] J. E. Nielsen, S. Thorndahl, and M. R. Rasmussen, "Improving weather radar precipitation estimates by combining two types of radars," *Atmospheric Research*, vol. 139, pp. 36–45, 2014.
- [34] A. Categorical, "Glossary of terms," *Machine Learning*, vol. 30, no. 2, pp. 271–274, 1998.