Paper:

Rain-Area Identification Using TRMM/TMI Data by Data Mining Approach

Shan-Tai Chen*, Chien-Chen Wu*, Wann-Jin Chen, and Jen-Chi Hu****

*Dept. of Computer Science, Chung Cheng Institute of Technology, National Defense University
No.190, Sanyuan 1st St., Tashi, Taoyuan, Taiwan
E-mail: stchen@ndu.edu.tw
**School of Defense Science, Chung Cheng Institute of Technology, National Defense University
[Received April 22, 2007; accepted September 20, 2007]

***Rain-area identification* distinguishes between rainy and non-rainy areas, which is the first step in some critical real-world problems, such as *rain intensity identification* and *rain-rate estimation*. We develop a data mining approach for oceanic rain-area identification during typhoon season, using microwave data from the Tropical Rainfall Measuring Mission (TRMM) satellite. Three schemes tailored for the problem are developed, namely (1) *association rule analysis* for uncovering the set of potential attributes relevant to the problem, (2) *three-phase outlier removal* for cleaning data and (3) the *neural committee classifier* (NCC) for achieving more accurate results. We created classification models from 1998-2004 TRMM Microwave Imager (TRMM-TMI) satellite data and used Automatic Rainfall and Meteorological Telemetry System (ARMTS) *rain gauge* data measurements to evaluate the model. Experimental results show that our approach achieves high accuracy for the rain-area identification problem. The classification accuracy of our approach, 96%, outperforms the 78.6%, 77.3%, 83.3% obtained by the *scattering index*, *threshold check*, and *rain flag* methods, respectively.**

Keywords: data mining, classification, rain-area identification, TRMM, microwave

1. Introduction

Heavy rainfall during the typhoon season leading to landslides, floods, and other weather-related disasters is a particular problem in and around the island of Taiwan, making it an important to calculate rainfall accurately to minimize potential weather- related damage.

Satellite observations provide such useful information as microwave channel data, visible image data, and infrared image data, for calculating surface rainfall, providing more frequent observations over a wider area than that of conventional *rain gauge* measurements. Limitations associated with such data, however, include the fact that infrared radiometers on satellites sense radiation emitted by cloud top, the inability of infrared observation to provide information on vertical cloud structure, and satellite image visible only during the daytime. Passive microwave radiometers, in contrast, can physically sense particles and hydrometeors within precipitating clouds. Therefore, microwave channel data are widely used in calculating heavy rainfall.

By means of these satellite-observed data, algorithms for precipitation problems are categorized as statistical, physical, or semi-physical. Due to the incompleteness of such information, however, algorithms may lead to biased results, estimating rainfall inaccurately. *Data mining* techniques in contrast are robust in determining patterns, associations, anomalies, and statistically significant structures in data intelligently and semi-automatically.

Rain-area identification is the first step in certain critical real-world problems, such as *rain intensity identification* and *rainfall estimation*, attracting much research, e.g., Cheng [1] used geostationary operational environmental satellite (GOES) infrared (IR) images for rain-area classification in Florida, attaining a classification accuracy of 63%. Gonzalez [2] used infrared and near-infrared AVHRR satellite data simultaneously to determine rainfall areas at accuracy from 75% to 91%, depending on the data sets tested.

The Tropical Rainfall Measuring Mission (TRMM), an international project jointly sponsored by the Japan Aerospace Exploration Agency (JAXA) and the United States National Aeronautics and Space Administration (NASA), uses the TRMM and TRMM Microwave Imager (TMI) satellite to obtained passive microwave and infrared data. Applications [3–5] and precipitation-related research [6–9] using TRMM/TMI data is widely used, and several types of rain-area identification have been developed using TRMM/TMI data, including the *Scattering Index* [10], *Threshold Check* [11], and *Rain Flag* [12].

Data used in this study was obtained from the TRMM/TMI satellite, a nine-channel passive radiometer with dual vertical and horizontal polarization microwave channels at 10.7, 19.4, 37, and 85.5 GHz and a vertical polarization microwave channel at 21.3 GHz. We used TMI level-1B11 brightness temperatures to recognize rain areas, and TRMM/TMI data from 1998 to 2005 to create classification models. We then used corresponding *rain gauge* data from Automatic Rainfall and Meteorological

Table 1. Nine TRMM/TMI microwave channels [14].
V: Vertical. H: Horizontal.

Channel	1	2	3	4	5	6	7	8	9
Frequency (GHz)	10.65	10.65	19.4	19.4	21.3	37	37	85.5	85.5
Polarization	V	H	V	H	V	V	H	V	H

Telemetry System (ARMTS) measurements to evaluate these models. We use 80% of data for model training and cross-validation, 20% for testing.

In the data mining approach we developed for oceanic rain-area identification during the typhoon season, we tailored three schemes to make classification more accurate: (1) *association rule analysis* for uncovering the set of potential attributes relevant to the problem, (2) *three-phase outlier removal* for cleaning data, and (3) the *neural committee classifier* (NCC). Experimental results demonstrated that our proposed model outperforms previous methods. We used this approach to recognize rain areas for the Mindulle typhoon that killed over a dozen people in Taiwan in 2004. Using only five data attributes and less computation time, our classification results agreed well with those obtained by the Goddard Profiling (GPROF) algorithm [13], a current rainfall estimation algorithm.

This paper is organized as follows: Section 2 introduces rain-area identification method. Section 3 discusses the *neural committee classifier*. Section 4 details our data-mining framework. Section 5 compares results for different methods, and Section 6 presents conclusions.

2. Background

We begin by reviewing rain-area identification algorithms using nine TMI microwave channels (**Table 1**).

2.1. Scattering Index

The Scattering Index (SI) [10] was developed for SSM/I microwave multi-channel data to recognize rain over the ocean through variations in brightness temperature (Tb). The SI is calculated using three vertically polarized radiances at 19.4, 21.3, and 85.5 GHz (Eq. (1)). In global application, an SI exceeding 10K indicates that the rain occurs at 1 mm/hr or more.

$$\mathrm{SI} = \mathrm{A}_1 + \mathrm{A}_2(\mathrm{Tb}_{19.4\mathrm{V}}) + \mathrm{A}_3(\mathrm{Tb}_{21.3\mathrm{V}}) + \mathrm{A}_4(\mathrm{Tb}_{21.3\mathrm{V}})^2 - \mathrm{Tb}_{85.5\mathrm{V}} \quad . \; . \; . \; . \; . \; . \; . \quad (1)$$

where $\mathrm{A}_1 = 450.20, \mathrm{A}_2 = -0.5060, \mathrm{A}_3 = -1.8740$, and $\mathrm{A}_4 = 0.00637$; Tb denotes brightness temperature (K); and the Tb subscript indicates the SSM/I channel.

2.2. Threshold Check

The threshold Check (TC) [11], developed by Chen and Li, calculates rainfall occurrence during the Mei-Yu summer monsoons season in Taiwan. The method derives a threshold by calculating the average Tb of TMI dual-polarized radiances from 10.65 to 21.3 GHz for all no-rain events. A rain scene is identified when the Tb of any TMI pixel exceeds the threshold.

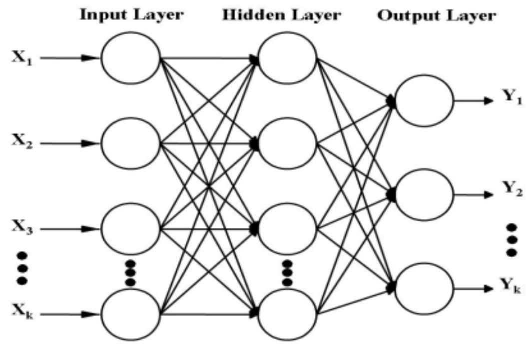

Fig. 1. BPN Architecture.

2.3. Rain Flag

The rain flag (RF) [12], developed by Goodberlet et al., to retrieve the oceanic wind speed in microwave observations, is divided by Tbs variations of 19H, 37V, and 37H GHz into four types, i.e., RF = 0, 1, 2, and 3, denoting *no rain*, *uncertainty, rain*, and *heavy rain*.

Classification rules and parameter settings we used for the three rain-area identification algorithms are as follows:

- SI: A TMI pixel is classified as raining if its SI exceeds 10K and Tb22V exceeds $38 + 0.88 \times$ Tb19V [10]; otherwise, it is classified as no rain.
- TC: A TMI pixel is classified as raining if the average Tb of 10.65V, 10.65H, 19.4V, 19.4H and 21.3V exceeds the threshold set by all no-rain events of the training dataset; otherwise, it is classified as no rain.
- RF: A TMI pixel is classified as raining if its RF = 2 or 3, and no-rain if its RF=0. Pixels with RF=1 cannot be classified by RF, so we omit these pixels in our experiments.

3. Neural Committee Classifier

The Neural Committee Classifier (NCC), a committee classifier for rain-area identification, consists of several neural network base classifiers. The artificial neural network (ANN) involves parallel processing that simulates biological neural systems. Two features make ANN a popular, powerful classification tool [15]: i.e., (1) it learns rules from training data and (2) it provides highly parallel computing power and fault tolerance. The back-propagation network (BPN) used by the NCC is a multi-layer feed-forward network containing input, hidden, and output layers (**Fig. 1**).