Paper:

A Data Mining Approach to Rainfall Intensity Classification Using TRMM/TMI Data

Shan-Tai Chen*, **Shung-Lin Dou***, **and Wann-Jin Chen****

*Dept. of Computer Science, Chung Cheng Institute of Technology, National Defense University
No.190, Sanyuan 1st St., Tashi, Taoyuan, Taiwan, R.O.C.
E-mail: stchen@ccit.edu.tw
**Dept. of Environment Information Engineering, Chung Cheng Institute of Technology, National Defense University
[Received June 7, 2008; accepted August 20, 2008]

The systematic approach we propose for classifying oceanic rainfall intensity during the typhoon season consists of two major steps – 1) identifying the rain areas and 2) classifying rainfall intensity into *normal* and *heavy* for these areas. The heterogeneous hierarchical classifier (HHC), an ensemble model we developed for accurately identifying heavy rainfall events, consists of a set of base classifiers. The base classifiers are independently constructed through heterogeneous data mining approaches such as artificial neural networks, decision trees, and self-organizing maps.
The meteorological satellite Tropical Rainfall Measuring Mission (TRMM) microwave imager (TMI) data from 2000 to 2005 are used to create the classification models. TRMM precipitation radar (PR) data and rain gauge data from Automatic Rainfall and Meteorological Telemetry System (ARMTS) measurement are used as ground truth data to evaluate models. Two thirds of the dataset is used for model training and one third for testing. Experimental results show that the proposed model classifies rainfall intensity highly accurately and outperforms previously published methods.

Keywords: data mining, classification, rainfall intensity, TRMM, microwave

1. Introduction

Heavy typhoon season rainfall causing landslides, floods, and other weather-related disasters is a critical problem in Taiwan, among other areas, making it important to predict areas of heavy rainfall to minimize potential weather-related disasters.

Among the extensive heavy-rainfall research done, Alexiuk [1] used meteorological volumetric radar data to detect thunderstorms and hybrid strategies to classify storm events into wind, heavy rain, tornados, and hail. Using cluster analysis, Martinez [2] classified heavy-rain events into eight atmospheric patterns and further determined the relationship between heavy-rain regions and the atmospheric patterns. Using self-organizing maps (SOMs), Nishiyama [3] identified a typical synoptic pattern causing heavy rainfall during the rainy season.

Satellite observations provide useful information, e.g., visible image data, infrared image data, and microwave channel data, for estimating surface rainfall. They also provide more frequent observations than conventional rain gauge measurement and cover a wider area than radar data. Lee [4] proposed pattern recognition using visible and infrared satellite images to classify rain rates into *none*, *light*, and *heavy*, achieving an accuracy exceeding 70%. Using visible and infrared samples from VHRR images, Parvathi [5] took the maximum likelihood decision rule and classified rain rate events into four categories for an accuracy exceeding 80%. Limitations associated with such data, however, are that 1) visible satellite images are available only during daytime and 2) infrared observations cannot provide information on vertical cloud structure.

Passive microwave radiometers physically sense raindrops and hydrometeors within precipitating clouds, so microwave channel data is widely used in heavy oceanic rainfall estimation. The TRMM/TMI [6] satellite provides passive microwave and infrared data for monitoring and studying tropical rainfall, and has been used in different precipitation-related applications [7, 8]. Several types of rain-area identification have been developed using TRMM/TMI data, including a scattering index [9], threshold check [10], and rain flag [11]. The rain flag is also used to identify heavy-rain events, classifying weather systems into four categories – no rain, uncertain, rain, and heavy rain.

In this paper, a new algorithm is developed for determining rainfall intensity during the typhoon season using TRMM/TMI data. As in [4], we classifies events into one of three rain-rate categories – none, normal, and heavy. We constructed two ensemble models – the neural committee classifier (NCC) [12] and the heterogeneous hierarchical classifier (HHC) – for rain-area identification and rainfall intensity retrieval. Experimental results show that our algorithm is highly accurate. Our model has been applied to determining rainfall intensity for an actual typhoon. Results show relatively high agreement between the Goddard profiling algorithm (GPROF) [13] and ours.

This paper is organized as follows: Section 2 presents

Table 1. Nine TRMM/TMI microwave channels [15].
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

a brief background and Section 3 details the HHC concept and the algorithm for constructing it. Section 4 provides the framework for our approach to rain-rate classification and Section 5 reviews experimental results. Section 6 gives conclusions.

2. Background

We begin by reviewing the background to this work.

2.1. Data Source

Datasets used for training and testing in this research include TRMM/TMI, TRMM/PR, and rain gauge data, which are introduced as follows:

TRMM Data

The international Tropical Rainfall Measuring Mission (TRMM) [6], jointly sponsored by the US National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA), provides passive microwaves with nine channels (**Table 1**), and infrared data for monitoring and studying tropical rainfall. The TRMM satellite, with a period of 91.5 minutes and launched in November 1997, provides satellite observations from 40°N to 40°S. TMI level-1B11 brightness temperature (Tb) data and TRMM/PR rainfall estimation data we have used here are downloaded from the NASA website [14].

Island rain gauge data

Rain gauge data used in this study is from the Japan Meteorological Agency (JMA) and the data range period is from July to October 1998-2005. **Table 2** shows the 11 islands providing rain gauge data (mm/10-min) we used here. Rain gauge data, with 60-minute accumulation rainfall, is used as ground truth for evaluating results. The data is accumulated from 30 minutes before and after the time when the satellite passed over a rain gauge [15–17].

We define a TMI pixel as a rain area if accumulated rainfall exceeds 1 mm/hr and as a heavy-rain area if it exceeds 15 mm/hr.

2.2. Rain Flag Algorithm

The rain flag (RF) [11], developed by Goodberlet et al., retrieves oceanic wind speed through microwave observation. Depending on variations in the Tbs of 19H, 37V, and 37H GHz, weather systems are classified into four types, i.e., RF = 0, 1, 2, and 3, which denote the *no-rain*, *uncertain*, *rain*, and *heavy rain* types, respectively. Classification rules for the four are as follows:

$$0 : \text{Tb37V} - \text{Tb37H} > 50\text{K and Tb19.4H} < 165\text{K} \quad (1)$$

$$1 : 37\text{K} \leq \text{Tb37V} - \text{Tb37H} < 50\text{K or Tb19.4H} \geq 165\text{K} \quad (2)$$

$$2 : 30\text{K} \leq \text{Tb37V} - \text{Tb37H} < 37\text{K} \quad (3)$$

$$3 : \text{Tb37V} - \text{Tb37H} < 30\text{K}. \quad (4)$$

We classify a TMI pixel as no rain for RF=0, normal for RF=2, or heavy for RF=3. Pixels classified as RF=1 cannot be classified by RF, so we omit them from our experiments.

Table 2. Island rain gauge station sites [15].

NO	Station	Code of Station	North Lat., East Lon.	Altitude(m)
1.	IRABU	93011	24.82°N, 125.17°E	10
2.	MIYAKOJIMA	93041	24.79°N, 125.27°E	40
3.	GUSUKUBE	93051	24.74°N, 125.41°E	55
4.	TARAMA	93061	24.66°N, 124.69°E	16
5.	IBARUMA	94001	24.50°N, 124.28°E	15
6.	KABIRA	94036	24.46°N, 124.14°E	7
7.	YONAGUNIJIMA	94017	24.46°N, 123.01°E	30
8.	IRIOMOTEJIMA	94061	24.38°N, 123.74°E	9
9.	ISHIGAKIJIMA	94081	24.33°N, 124.16°E	6
10.	OOHARA	94101	24.26°N, 123.87°E	28
11.	HATERUMA	94116	24.05°N, 123.76°E	32

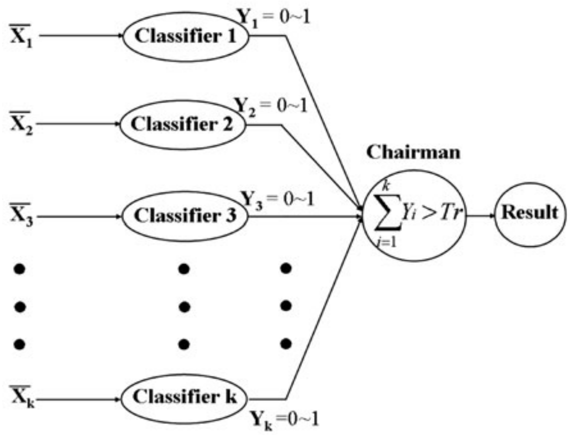

Fig. 1. A logical view of NCC [12]. It conducts classification by aggregating predictions made by individual weak classifiers.

2.3. Neural Committee Classifier

The neural committee classifier (NCC) [12], an ensemble classifier for rain-area identification, consists of several weak classifiers (WC) of neural networks (**Fig. 1**). X_i and Y_i are the input vector and output value for each WC i. Each WC is a multilayer neural network trained independently by the back propagation algorithm using different TRMM/TMI microwave channel data. The chairman function, which judges whether a TMI pixel is a rain area