Term Project: Detection of Small Craters in High Resolution Planetary Images Using Shape and Texture Features

Assigned Date: Thursday, March 25, 2010

Educational Goal

Apply spatial data mining techniques to real-world applications.

Phase I: Training Set Analysis

(100 points)

Due: 8:30 PM Thursday April 1, 2010

Requirements

- Training data is in CSV format (Comma-Separated Values) . Download the file from Term Project Phase I: information package via UMass Online.

  Note that you cannot open the CSV file using MS Excel Spreadsheet because the data would be crashed if you do so. You should read the file using a programming language, for example, Java or Matlab.

  Training set description:

  There are 496 crater candidates in the training set (Rows = 496), each candidate has 1093 attributes (Columns = 1093). There are 204 true craters and 292 non-craters in this training set. The 1st and 2nd columns are the x and y coordinates of the centroid of a crater candidate (measurement unit is pixel); the 3rd column is the radius of the crater candidate (measurement unit is pixel); columns 4 – 1092 are the 1089 texture-based features of the crater candidate (data type is double); column 1093 is the class label, where 1 is for crater and 0 is for non-crater.

- Convert the training set from CSV format to Weka ARFF format. Implement a script for the data conversion, using a language of your choice.
- Use Weka J48 to build a decision tree from the training set using 10-fold cross-validation. Design 20 experiments to gradually build Underfitting decision trees to overfitting decision trees. Report accuracy, error rate, precision, recall, and F-measure for the 20 experiments. Implement a script to call Weka J48 API to run the 20 experiments and automatically extract the experiments results on accuracy, error rate, precision, recall, and F-measure.

Submission Requirements

1. Write an experiment report to discuss the 20 experiments, including parameter settings and experimental results. Submit the paper copy of the report, and source code of the scripts with the cover page in class. Paper copy should be bound firmly together as one pack (for example, staple, but not limited to, at the left corner). 5 points will be deducted for unbounded homework.
2. Submit the softcopy of the report and scripts through your UMassOnline account at http://boston.umassonline.net/index.cfm.
3. Zip all the files. One submission per team. Save the file as sdm_ teamNumber. For example, Team 1 should name their file as *sdm_team1.zip.*
4. No hard copies or soft copies results in 0 points.