Programming Assignment 3

(200 points)

Assigned Date: Wednesday, November 30, 2011

Due Date:

Phase I Pseudocode presentation	**5:30 PM Monday, December 7, 2011**
Phase II Implementation & Presentation	**6:30 PM Monday, December 19, 2011**

Educational Goal

Applying AI algorithms to a real world problem of crater detection from planetary images.

Requirements

- **Training set and Test sets**. The crater dataset is described in http://kdl.cs.umb.edu/w/henryzlo/2011/11/30/crater-dataset/.
- Both training set and test set are labeled. "Positive" is for craters and "Negative" is for non-craters. Note that you should not open the CSV files directly using MS Excel Spreadsheet because the data would be crashed if you do so. Read the files using a programming language, for example, Java or Matlab.
- Design your own algorithms or adapt existing feature selection and classification algorithms to learn a classification model from the training set and classify the test set. You cannot use class label information in the test set and you can only use the class label information in the test set for external evaluation. On the other hand, you should take full advantage of the class label information in the training set. You may consider using Weka, data mining software in Java, in this assignment instead of implementing your code from scratch. You may use or change any existing algorithms or tools in this assignment, but you must justify your own contribution in the final report.
- **Benchmark Dataset for Evaluation.** Report the best results of your learning algorithm in **precision, recall, and F1 scores** using the complete test set. The complete test set consists of the Positive Test Set and Negative Test Set posted at http://www.cs.umb.edu/~henryzlo/.
- (50 points) Phase I: Prepare a 5-minute presentation to discuss your initial design of pseudocode of the proposed strategies.
- (150 points) Phase II: Design and implement the software and report your best results. Prepare a 20-minute presentation in class to discuss the proposed method and experimental results.

Submission Requirements

1. Submit the one-page PPT slide and your project files and results on December 7, 2011 and December 19, 2011, respectively.
2. Your program should be well-documented. Variable names and function names should be self-descriptive. Major functions should be explained clearly in comments. The program outputs should be presented in a clear sequence.
3. Turn in the paper copy **including the results** and soft copy of all the files of this assignment. Submit a single zipped file of all the files of this assignment through your UMassOnline account at http://boston.umassonline.net/index.cfm. Submit the paper copy along 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.
4. Name your file with AI_ lastname_firstname_pr3. For example, student John Smith should name his file as AI_Smith_John_pr3_phaseI and AI_Smith_John_pr3_phaseII.
5. No hard copies or soft copies results in 0 points.