Homework Assignment 4

(200 points)

Assigned Date: Tuesday, April 23, 2013

Due Date of Progress Report and Final Report:

Progress Report 1: 4:00PM Tuesday, April 30, 2013

Progress Report 2: 4:00PM Tuesday, May 7, 2013

Final Report: 4:00PM Tuesday, May 14, 2013
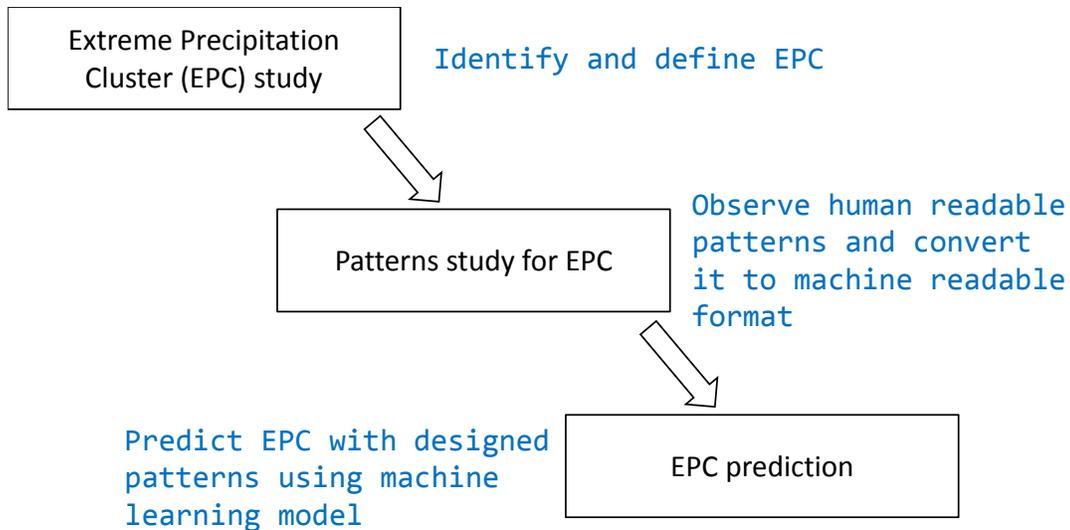
Educational Goal

Apply machine learning techniques to long-lead forecasting of extreme flood events; Perform interdisciplinary teamwork between CS students and EEOS students.

Project Requirements

1. **Flood data**: Contact your TA Yang Mu to obtain the 23,011 observations over 63 years from January 1st, 1948 to December 31st, 2010. The data set includes 9 meteorological variables described in Homework 3 as well as in Table 1 of the paper handed out (Referred as the Flooding Paper in the rest of the assignment description) in class.

   > **Need additional information?** Please talk to your professor Wei Ding and we will contact our domain scientists to obtain the information for you.

2. **Project Topic**: Every team must improve class label definition in Section 4.1 Class Label of the Flooding Paper. You may design a totally different or better way to define class label of extreme precipitation clusters (EPC).

- **Your own idea on flood prediction**: You can propose your project idea and report your preliminary results in Progress Report 1 due on April 20; or

- **Suggested topic 1**: Improve the method of feature construction in Section 4.3 Feature Space Construction of the Flooding paper and then perform flood prediction afterward. Please find a better or different way to construct spatial and temporal meteorological variables.

- **Suggested topic 2**: Improve training set construction in Section 5.3 Hierarchical Re-sampling and then perform flood prediction afterward. Please find a better or different way to deal with imbalance training set.

- **Suggested topic 3**:

Extreme Precipitation
Cluster (EPC) study

Identify and define EPC

Patterns study for EPC

Observe human readable
patterns and convert
it to machine readable
format

Predict EPC with designed
patterns using machine
learning model

EPC prediction

Three steps are as follows.

1). Extreme Precipitation Cluster (EPC) study

We hypothesize that flood is caused by Precipitation Cluster (PC). Therefore, we target at finding EPCs (Extreme Precipitation Cluster) from PCs because EPC is hypothetically associated with extreme flood events.

Open research problem: provide a formal definition of EPC and PC.

2). EPC Pattern study

After we formally define EPC and PC in Step 1, we can identify patterns for PCs (this is what you have done in Homework 3). Our goal is to find 2 sets of patterns: a positive set (patterns always occur in both PCs and EPCs) and a negative set (patterns are for PCs but not for EPCs).

3). EPC prediction

Design a machine learning algorithm to use the positive and negative sets identified in Step 2 for EPC prediction.

Report Requirements

1. Your reports must clearly
   - Explain the motivation of the proposed method.
   - Discuss why the proposed method is better.
   - Describe how to quantitatively analyze the proposed method.
   - Discuss how to design the experiments to evaluate the proposed method.
   - Discuss how to validate the proposed method using various experiments.
   - Illustrate the experimental results using figures and/or tables.

Submission Requirements

1. Attach the homework cover page to your report.
2. Submit the softcopy of your report and code via UMassOnline. Zero points for late submission.
3. Turn in the paper copy of your report 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. No hard copies or soft copies results in 0 points.