EE 660 Project Proposal Form Fall 2015

Instructions. For full credit in HW 8 Problem 2, you will need substantive and reasonable answers to at least 4 of the fields with an asterisk (*). For the project proposal, you will need to answer all parts except the last field ("other comments") which is optional. Preferred format is to enter your answers into the Word version of this form, then convert to pdf before submission. If you don't like to use Word, then submit a typed version with each field labeled with its title ("Dataset used", etc.).

Project Title
Your name and email address
Xuechao Wu Xuechaow@usc.edu
If a team project, state how many team members (including yourself) and list them
Myself
***Dataset Used**
West Nile Virus Prediction The training set consists of data from 2007, 2009, 2011, and 2013, while in the test set the task is to predict the test results for 2008, 2010, 2012, and 2014. Specifically, the data set consists of three data sets. Mean data set has social and geographical data from traps and satellite traps. Two support set are weather set and mosquito spray set. They are significant side-effect on the main trap results.
***Software packages, language, and code**
I This research will use Matlab as primary tool and pmtk along with its toolboxes. Opecncv and C++ are secondary support tools. Matlab can conveniently handle the numerical data and string data, and the tool box pmtk is strong and powerful. Use Github for version controll
***Clear statement of the problem and/or goals.**
The problem is to predict the test results for 2008, 2010, 2012, and 2014, given training test as data from 2007, 2009, 2011, and 2013. Specifically, the Boolean variable of data "WNVpresence". Need to finish a calibrated model

using primary and preprocessed data. Need an result of error rate on successful predicts. Need a decision process with confidence interval.

***A plan of preprocessing and feature extraction**

Name	Description	Usable
ID	**The id of the record**	**Yes**
Date	**Date the test is performed**	**Yes**
Speicies	**The species of mosquitos**	**No**
Trap ID	**The id of the trap**	**Yes**
NumMosquitos	**The number of the mosquitos**	**Yes**
Latitude Longitutde	**Geography data**	**Yes**
Block Address Street address accuracy	**Social location description**	**No**
WNVpresented	**Whether the virus presents**	**yes**
Weather data and spray location	**Additional data**	**no**

Virus is contagious so location is very important in this data set. So one primary task is to conclude a cluster information with raw data of latitude and longitude, by address information with a weight vector using accuracy data.

By the spray information and weather condition including temperatures and humidity to generate a mosquito activity strength model. Specifically, the model takes temperatures, humidity spray and location (or cluster division results) as input, species and amount as to output.

Besides the specific extraction above, further information might be needed: temporal change of mosquito activity, like the periodic regulations. And spatial information: the movement of mosquito group.

***A plan of your approach**

1. methods to try for classification or regression
   Basically, the virus's propaganda relies on the mosquito group movement, weather and anti-bio spray condition. So the proper way is to use regression tree.

2. models and hypothesis sets
   First, I will use full linear weight vector. Afterward, add lambda, or weight decay into the model. There are cross relations on mosquito group movement , weather, and spray work. Establish three individual models for their conditions and the output of WNVpresence. Use another logistic regress to combine these three model together into a general prediction model.
3. How you will use datasets (test sets, training sets, etc.)
   For training sets, use cross validation to assure the precision. Use test sets to verify Eout.
4. how you will evaluate your system's performance
   The error rate of prediction of next year.
5. A timeline can also be useful.
   Week1: Data pre-processing
   Week2: Model Computation
   Week3: Error Test and Model Revision
   Week4: Paper writing

***A description of any prior or parallel work of yours**

None

***If yours is a team project, roughly describe how work will be divided**

No Team

Other Comments

The proposal is still under revision. Further information will be added with the project going on.