CSE 515 Multimedia and Web Databases

Phase #1

**Group Members Advisor**

Nagarjuna Myla **K. Selcuk Candan**

Siddharth Sujir Mohan

Jacob Mims

Tsung-yen Yu

**Abstract**

In this project we did experiment on epidemic simulation data sets and time series. Created an epidemic words dictionary, using it identified epidemic words average and differences of the 1 hop neighboring states. This generated data is used in identifying the highest and lowest strengths and their one hop neighbor states in the given data set and highlight them on heat map of all the states across time series for the given simulation file.

**Keywords**

Simulation, epidemic, strength, highest, lowest, data sets, time series, 1 hop, neighbors, states, connectivity graph.

**Introduction**

**Terminology**

dir: directory where the simulation data sets located

w: window length which indicates the number of samples on a time series

h: shift length which indicates what length we need to shift window in each iteration

r: resolution which indicates the number of Gaussian bands into which the data from data sets to be distributed

: mean

: standard deviation

s: state

: state i

t: time

f:file

idx:<f,s,t> pair

: quantized vector of window length

: length of band i which is derived by calculated using Gaussian distribution.

epidemic\_word\_file: epidemic words dictionary

G: Connectivity graph

: weight

=<,>

=<,>

=<,>

1HN(G,s): 1 hop neighbor of s as denoted by G

: denotes element wise division operation

norm-2 ||: strength of word

**Goal Description**

For a given data sets of epidemic simulation files generate heat map visualizing the highest and lowest strength states and their 1 hop neighbor states in connectivity graph G.

**Assumptions**

* User will provide the directory where simulation files are present.
* User will provide the directory where output files will be stored.
* User will provide the directory where LocationMatric.csv is located (Connectivity Graph).
* User will provide required data when prompted for r, w, h, .
* Maintained separated epidemic\_word\_file, epidemic\_word\_file\_avg,epidemic\_word\_file\_diff for each data set to improve efficiency for algorithm.
* On the heat map pink colored rectangles indicate lowest strength state (which is having an ellipse) and 1hop neighbors of it.
* On the heat map white colored rectangles indicate highest strength state (which is having an ellipse) and 1hop neighbors of it.

**Description of the proposed solution/implementation**

To achieve the goal we go through three tasks, output of each task is used as input to other t asks details on each task are explained below.

**Interface specifications**

Task1:

1. Normalized the data sets in each file between 0.0 to 1.0
2. Quantized each of the value into r Gaussian bands with parameters =0 and =0.25 by calculating length of each band i using below formula and center of band is taken as representative of that cluster.

=

1. For each state a window w-length is moved on time series and noted for each iteration i.

Task2:

For a given G, epidemic\_word\_file and alpha we do calculate and for each using below formula.

= () + ((1-|1HN(G,.s))^ =<.f,,.t>)})

= (|1HN(G,.s))^ =<.f,,.t>)}))

Task3:

For each word calculated norm-2 ( strength) on each epidemic\_word\_file, epidemic\_word\_file\_avg, epdidemic\_word\_file\_diff file and found highest and lowest strength. Given simulation file is shown as heat map and corresponding highest and lowest strengths along with 1HN(G,S) on the selected word file.

**System requirements/installation and execution instructions**

1. Matlab R2013 or higher
2. RAM size > 1 GB
3. Execute Task1, Task2, Task3 sequentially.

**Related work**

**Conclusions**

**Bibliography**

**Appendix**

This project is worked individually by each of group member. I completed this task and helped my group members in their doubts on how to implement parts of the tasks when they got stuck. I gave directions on how to improve efficiency as program execution time is taking long for them. With this project I got a chance to learn about a tool Matlab which I never worked or used. I learned myself went through tutorials on Matlab and now I got a complete overview on Matlab which I found so hard for the first two days of project start and this tool made the work ease.