**Work summary**

**Songnan Wu**

**04/26/2018**

**Analyzing path enumerator**

Given an enumerated path solution on a 5x5 maze grid, analyze the clusters(an area that formed by connected cells that are not passed by the solution path) in the maze generated by each path, and fetch the following information for each single solution path:

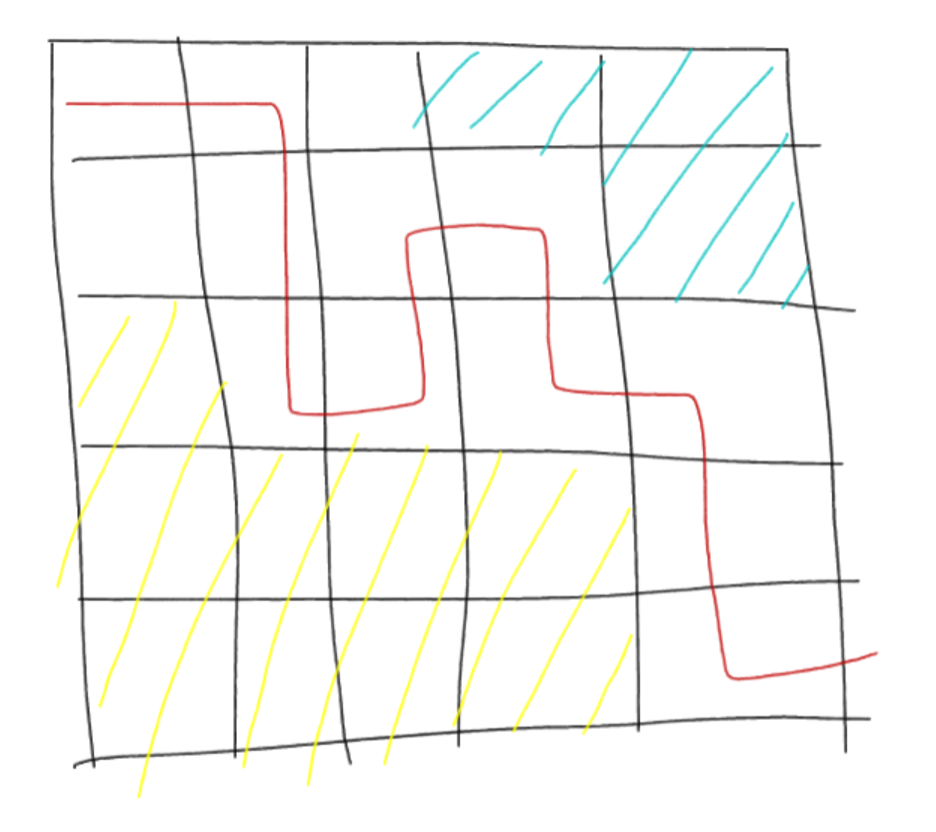
1. Number of clusters in the maze map generated by a solution path
2. The cluster with max size (number of cells per cluster) among all clusters generated by one solution path
3. The cluster with min size (number of cells per cluster) among all clusters generated by one solution path
4. The median size number of all clusters generated by one solution path
5. The cluster with largest surface area (the surrounding cell sides that forming the cluster) among all clusters generated by one solution path
6. The cluster with smallest surface area (the perimeter of the cluster) among all clusters generated by one solution path
7. The median number of surface area all clusters generated by one solution path
8. The cluster with largest compactness (cluster size/bounding box size) among all clusters generated by one solution path
9. The cluster with smallest compactness (cluster size/bounding box size) among all clusters generated by one solution path
10. The median number of surface area all clusters generated by one solution path
11. The path length

After generating all the info for a path, a dictionary will be created to store all the info as a set of keys, and the value will be the amount of path lengths with all 11 elements in the key set possessing the same data.

**Data Output Example**

For example, for a given solution path like this:



red line: the solution path from top left to bottom right, every solution path will start from top left cell and goes to the bottom right cell

shadows: the clusters generated based on the path

For cluster 1, the cluster size will be 3 since it’s combined with 3 cells; the surface area is 8, is the sum of the total sides that form the cluster, could be understood as perimeter. The compactness is ¾, where 3 is the cluster size and 4 is the bounding box size.

For cluster 2, the cluster size will be 9 since it’s combined with 9 cells; the surface area is 14, is the sum of the total sides that form the cluster, could be understood as perimeter. The compactness is 9/12, where 9 is the cluster size and 12 is the bounding box size.

So, based on this solution path,

1. Number of clusters: 2
2. Max cluster size: 9
3. Min cluster size: 3
4. Median cluster size: 6
5. Max surface area: 14
6. Min surface area: 8
7. Median surface area: 11
8. Max compactness: 750 (0.75 multiplied by 1000)
9. Min compactness: 750
10. Median compactness: 750
11. Path length:11

**User-defined path collections**

The program would ask users to input what features from above 11 information they want to specify (enter -1 if they don’t care about this feature), and the path vectors (shown in the form of cell numbers that the path go through) that meet the requirement would be outputted.

**Histogram**

1-D histogram:

For each feature (all 11 info above), generate a histogram with the number of paths for all different possible feature value. For example, for all the possible path lengths (could be 9, 10, 11, etc.), get the amount of paths for each different path length and accordingly generate the histogram, 11 in total.

2-D histogram:

Generate a 2-D histogram, set the x value to be the path length, and the y value to be any of the remaining 10 data, and the z value to be the amount of paths possessing the same x and y value characteristics.

**Output File**

ClusterInfo-PathAmount.txt: the file that stores all possible cluster info combinations and the amount of paths possessing each combination info.

AllPathVector.txt: all the possible path vectors in a 5x5 maze

1DHistogram folder: all the data file for generating 11 1D-Histogram charts.

2DHistogram folder: all the data file for generating 2D-Histogram charts.

1DHistogramRawData folder: all the data file for the according 2D-Histogram in a more readable way

Histogram.xlsx: the excel file stores the charts for both 1D-Histogram and 2D-Histogram charts.