Spatial Similarity Measure

GitHub repo:

<https://github.com/freindst/spacial-similarity-measure.git>

Source code is in Python. On the Windows machine, Fiona and GDAL have to be installed through local .whl files which are provided in the git. All diagrams and data in tables can be found in the Jupyter notebook.

Related Package/models:

* Shapely
* Similaritymeasures
* Nltk
* geopandas

# 1 Proposal

* Find a way with better performance in searching similar spatial shapes
* Use bit string generated from the bit map to search similarity
* Use multi-resolution sketch to generate a signature for the geometry to search similarity

# 2 Method

## 2.1 Target Metrics

Hausdorff distance and Fréchet distance are the target similarity indicators I want to use other measurements to predict.

## 2.2 convert 2-d bit map of geometry to bit string

### 2.2.1 Generating bit map from geometry

A bit map can represent geometry. 1 means the geometry intersects with the pixel, and 0 doesn’t.

A picture containing text, black, keyboard, electronics

Description automatically generated

Fig.1 Bit-map of a geometry

In order to compare the similarities, 2d bit map can be reduced to one dimensional array. One way is concatenate each row one by one, another one is to adopt Z-order pattern.

Converting to bit string may lose some locality information, and there are some other factors to be considered.

1. Resolution of the bit-map or the sample size
2. Shingle size of Jaccard distance.

### 2.2.2 Result

Chart, bar chart

Description automatically generated

Fig. 2 Line-by-line grid bit string hash & Jaccard-distance correlation

Chart, bar chart

Description automatically generated

Fig. 3 Z-wrap Grid bit string hash & Jaccard-distance correlation

In the experiment, it shows that the grid number, which represents the pixel size, should be a value not too large. 50- to 100-grid is a good setting. The best shingle size for Jaccard distance is 16, which is also neither too large nor too small. It may be due to the more locality information are lost when the grid number increases, and the shingle size reduces.

Chart, line chart, histogram

Description automatically generated

Fig. 4 Jaccard-distance vs Hausdorff Distance using line-by-line method

Chart, line chart, histogram

Description automatically generated

Fig. 5 Jaccard-distance vs Hausdorff Distance using z-curve method

The performance of calculating bit string Jaccard distance has no advantage to Hausdorff distance or Fretcher distance. It is possible we can tune this method for better predicting results, but it is not a potential solution.

## 2.3 Multi-resolution trajectory sketch

This method randomly scatters round screen masks on the map and documents the round masks which have intersections with the geometry trajectory. The list of screen masks can be used to represent the trajectory. Here is a graph demonstrating the principle:

A picture containing shape

Description automatically generated

Fig. 6 Round Screen Masks and polygon on the canvas

The list of screen masks’ id is used as the signature of the 2-d shape. By changing the density and size of the round mask, we can generate multiple versions of the signatures. By comparing multiple sets of signatures, it is possible to predict the similarities without matching through shape.

## The following are experiments to compare the Jaccard distance of multi-resolution sketch with Fretcher distance.

Chart, bar chart

Description automatically generated

Fig. 7 Shingle size = 5 Jaccard Distance correlation between Fretcher Distance

Based on the experiment, a single signature cannot predict Fretcher Distance well enough. Instead, use multiple signatures may reach the goal.

# 3 Prospect Work

It is suggested to develop Nearest Neighbor search using LSH for similarity search. After the sample screen masks are set, to get the signatures of geometry is simple. It can be saved statically with the shape geometry for future searches. Instead of comparing all geometry shapes Hausdorff distance or Fretcher distance, the multi-resolution trajectory sketch signature seems to have better future in applications. The next step is to determine how to get a constant signature, how many versions are enough and how to predict the similarities.