A Variation of Assignment Problem Model in Identifying Intersection Relations between Datasets

Wenjun Yang
Department of Geography & AMOS, University of Kansas

Introduction

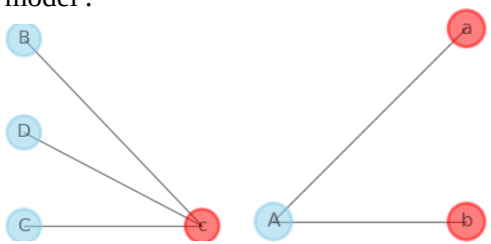
The technology of matching geospatial data plays a critical role in GIS and Remote Sensing when researchers try to harmonize datasets made by different agents or at various points of time. For example, the building footprints extracted from the same remote sensing image may differ because of the different agents or technologies.

Street view and unmatched building footprints between OSM and BingMap

Generalized Assignment Problem

To generate the optimal matching result, this study will introduce an optimization-based matching method. The model we refer to is one variation of the classic optimization problems: the generalized assignment problem. The model allows one worker to be assigned to multiple tasks. This concept reflects a "one-to-many" relation where one feature from one dataset has multiple corresponding features from the other datasets. Similarly, the model can also identify “many-to-one” relation by changing the position of two datasets in the model .

“Many-to-one” and “one-to-many” are two separate matching types in the present study. However, since the only difference is that their directions are opposite, we can use the same model to handle both scenarios.



One-to-many Many-to-one

Model

The objective is to minimize the total distance of all feature pairs. I and J are two datasets with m and n features. if the matching direction is from I to J, x_{ij} equals to 1 when i is matched to j, otherwise 0; c_{ij} is the distance of i of I and j of J; b_i represents how many feature j can be matched to i; y_i equals to 1 when i is selected, otherwise 0.

Objective:

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \quad (1)$$

Subject to:

$$\sum_{i=1}^{m} x_{ij} = 1 \quad j = 1...n \quad (2)$$

$$\sum_{j=1}^{n} x_{ij} \leq b_i * y_i \quad i = 1...m \quad (3)$$

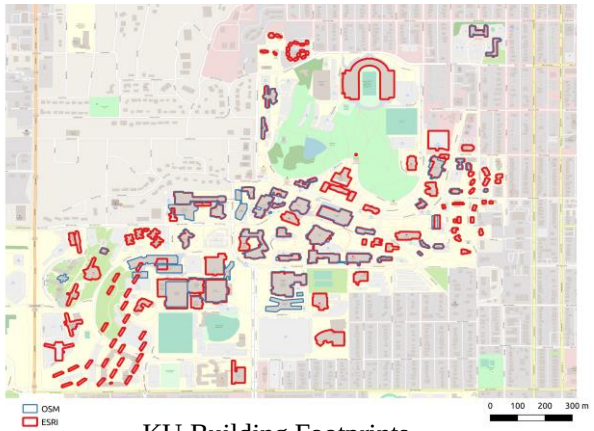
$$x_{ij} = 0 \quad or \quad 1 \quad (4)$$

$$y_i = 0 \quad or \quad 1 \quad (5)$$

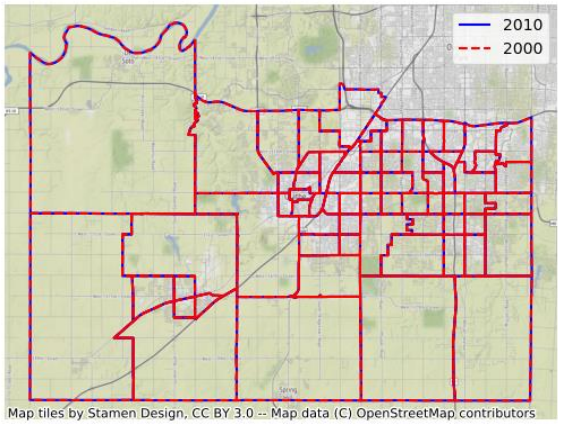
Generalized Assignment Problem

Application

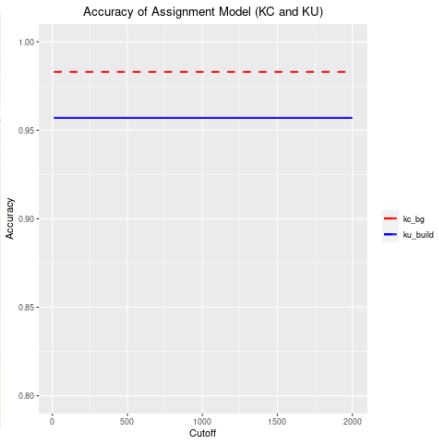
The matching method has been tested on matching datasets from different points of time and different mapping institutes. One example is to match KU building footprints from OpenStreetMap (OSM) and ESRI. Another example is to match part of census tracts of Kansas City Area in 2000 and 2010.



KU Building Footprints



Kansas City Census Tracts



Workflow

Data pre-processing
Set up a universal projected coordinate system.
Assign unique index for features.
Set proper threshold.

Measuring
Measure the similarity distance between all feature pairs using the Hausdorff distance.

Matching
Use optimization-based model to identify the optimal match plan.