M-SA-20230801

Normalized Inverse Distance with Marginal Utility Weighting (NIDMUW)

In the case of determining utility and preference value spatially

Kevin Sutjijadi @ 2023

Abstract

In the pursuit of finding a method to represent spatial properties, where multitudes of instances are also expressed, this method is developed, combining Inverse Distance Interpolation, Function Normalization, and Marginal Utility.

Summary

The function proposed for normalized inverse distance with marginal utility weighting (NIDMUW) is the following:

$$F(k) = \sum_{asc.\{j \in J, d_{jk} \leq D\}, n \leq N}^{n=0} W_j \cdot \left(1 - \frac{d_{jk}}{D}\right) \cdot e^{-\frac{\alpha \cdot d_{jk}}{D}} \cdot \mu^n$$

Equation 1

n = iteration number of features within reach
N = iteration number limit
j = feature
k = location/point of value
J = set of features
d_{jk} = distance from feature to location
D = search distance limit
Wj = weight of feature
α = exponent multiplier
μ = marginal multiplier
asc.= sort set ascending

Where the resulting value is the sum of all feature value within reach, sorted ascending by its distance (smaller to larger) and within the iteration number limit; Which feature value is the multiplication of feature's weight, normalized exponent function, and marginal multiplier with the power of the iteration number index. For inverse linear calculations, the exponent multiplier (α) can be set to 0, which cancels out the exponent.

The method intends to represent the preferential sway, or influence of a group of features to a particular location i.e. how good is a type of amenity coverage at that point. Application of this is the computation of discrete points or done in a matrix/array for raster data.

1. Introduction

IDW (Inverse Distance Weighting) has been one of the most common methods for assessing closeness/ reachability/ effect of certain features on given location, which could be any distance, point, or area (Franke 1982; Nalder and Wein 1998). It has been a staple on spatial related calculations and analysis, where providing a value that diminishes inverse to the distance from a feature has many applications and applicable on a number of cases (Shepard 1968; Franke 1982).

Modifications on this method are also common to further suit the context of each situation. As on this method, the aim is to provide an intuitive way of calculating a value that represents the preferential sway, or utility-ness on a given location based on its distance to multiple features in range. Where the method will require to incorporate function normalization (Miller 2006) to contain within 1 to 0 within a given distance limit to the distance factor; And also marginal multiplier, with a decreasing modifier for the subsequent in-reach features.

2. Things at a distance

The intended use of this method is for calculating the value of how much is a given location covered of a set of amenities and/or other features and expressing them as close as possible to emulate how it is perceived for its utilitarian value.

The intuitive way of perceiving how covered a location is from a set of amenities/features is how far it is from the given group of features, where the further it is, the smaller the distance, and with a distance limit can be imposed, as further than this distance, the feature has no influence on the location, rather than slowly creeping up to 0. Furthermore, multiples of in-range features does help to raise preference, but not equally, as subsequent further features will have a smaller influence due to redundancy as represented on Marginal Utility Concept.

From the basic IDW, all values are summed equally, or taken by the largest/closest one as stated as

$$F(k) = \sum_{i=2}^{n} \omega_i \cdot f_i \qquad \textit{Equation 2}$$

where f_i as the prescribed value of the dataset and w_i as the weighting component which is

$$\omega_i = \frac{d_i^{-p}}{\sum_{i=i}^{n} d_i^{-p}} \qquad \textit{Equation 3}$$

where the resulting value represents the average value by all distances, that builds up a continuous step value in between the values of the dataset.

However, for the purpose of this method, the desired output is not for interpolation but for determining a value based from a distance from the set value, with farther away from a given the feature results in a lower score.

Starting from the basic distance variable, the shorter the distance, the higher the value, while the longer, the smaller it gets. This function is commonly expressed inside Inverse Distance Weighting as

$$f(k,j) = \frac{1}{d_{jk}^{p}} \qquad \textit{Equation 4}$$

where it is simply one over the power (*p*) of a distance. Using this part results in possible values of infinite (at 0 distance), up to near 0 (at infinite distance), as well as changing values between different units of distances; As represented on Figure 1, where the red graph is where the power value is 1.0, and blue is 2.0.

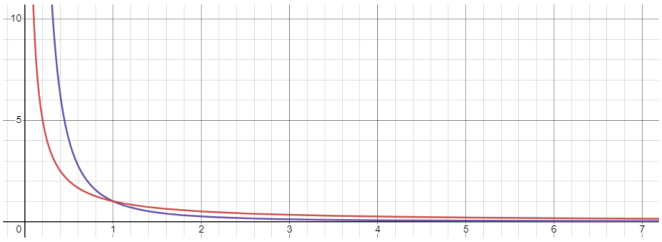


Figure 1 Graph of inverse distance

In the context of use, this may skew and distort the values and weighting (since a very small, weighted feature can result in an astronomically high value if the distance is near 0). To parametrize the resulting value to a normalized range, we could apply the common normalization function as

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad \textit{Equation 5}$$

and adjusted it with the context as to that at 0 distance, the value reaches 1 and at the limit distance, the value reaches 0. Moreover, the exponent part of the equation is also applied to this principle, resulting as

$$f(k,j) = \left(1 - \frac{d_{jk}}{D}\right) \cdot e^{-\frac{\alpha \cdot d_{jk}}{D}} \qquad \textit{Equation 6}$$

where d_{jk} is the distance from the feature to the location, and D is the distance limit, and α represents the exponent multiplication.

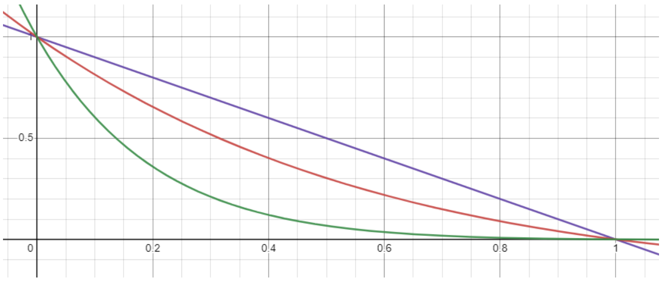


Figure 2 Graph for normalized inverse distance

As shown in Figure 2, where the blue line represents α = 0.0 for linear operations, and red line at α = 1.0, and green line at α = 4.0, which the higher it is, the steeper the drop. This enables further model fitting can use the exponent multiplier on how steep the distribution is. Furthermore, on all these plots, the D (distance limit) value is set to 1.0, to which all plots meet the x axis at x = 1. Which shows that it can adjust with the used index for the calculation as shown at Figure 3; Where red graph represents D = 1, blue at D = 10, green at D = 20.

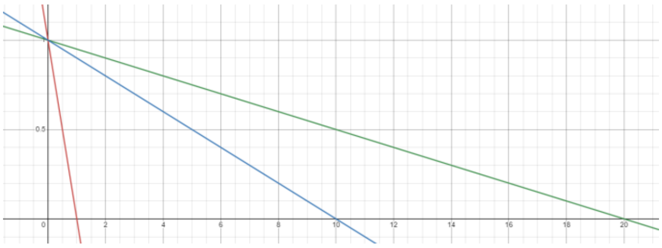


Figure 3 Graph demonstrating normalization on distance limit

Further elaboration on distance limit, this enables the calculation to be switch upon any form of units, either in meters, miles, kilometers, etc. can be directly compared as long as the distance and the distance limit is within one unit type.

for subsequent farther features, as their utility value would diminish (F. Y. Edgeworth, 1881). The addition of each value is modified by a multiplier in the power of its sorted index, ascending by distance, as

$$F(k) = \sum_{asc.\{j \in J, d_{jk} \leq D\}, n \leq N}^{n=0} f(d_{jk}) \cdot \mu^n$$

Equation 7

where μ is its constant multiplier that is in the power of the index of the list. This results in further, and therefore smaller weighted features will get marginal weighting as well.

Given all the factors and concepts in consideration, with the normalization of distance weighting and the marginalization of subsequent features, the resulting function is the following:

$$F(k) = \sum_{asc.\{j\in J, d_{jk}\leq D\}, n\leq N}^{n=0} W_j \cdot \left(1 - \frac{d_{jk}}{D}\right) \cdot e^{-\frac{\alpha \cdot d_{jk}}{D}} \cdot \mu^n$$

Equation 8

n = iteration number of features within reach
N = iteration number limit
j = feature
k = location/point of value
J = set of features
d_{jk} = distance from feature to location
D = search distance limit
Wj = weight of feature
α = exponent multiplier
μ = marginal multiplier
asc.= sort set ascending

Where the resulting value is the sum of all feature value within reach, sorted ascending by its distance (smaller to larger) and within the iteration number limit; Which feature value is the multiplication of feature's weight, normalized exponent function, and marginal multiplier with the power of the iteration number index. For inverse linear calculations, the exponent multiplier (α) can be set to 0, which cancels out the exponent.

3. Computation and application

Following the complex nature of the calculation, where there are limits, conditions, sorting, and dynamic factors, a corresponding computation function is developed; as this method will be applied as an interpolation method in spatial arrays, as well as direct input of certain coordinates.

The language used for this method is Python with addition of the NumPy, and for raster/spatial array calculations, GDAL libraries. For graph illustrations, SciPy and Matplotlib are also included. The sample dataset comes from an OSM (Open Street Map) request, consisting of health facilities in Jakarta. The bounds of the sample are:

EPSG = 32748 (WGS 84 48s)

Geotransform = (692180, 10, 0, 9323840.0, 0, -10)

Array Size = (800, 800)

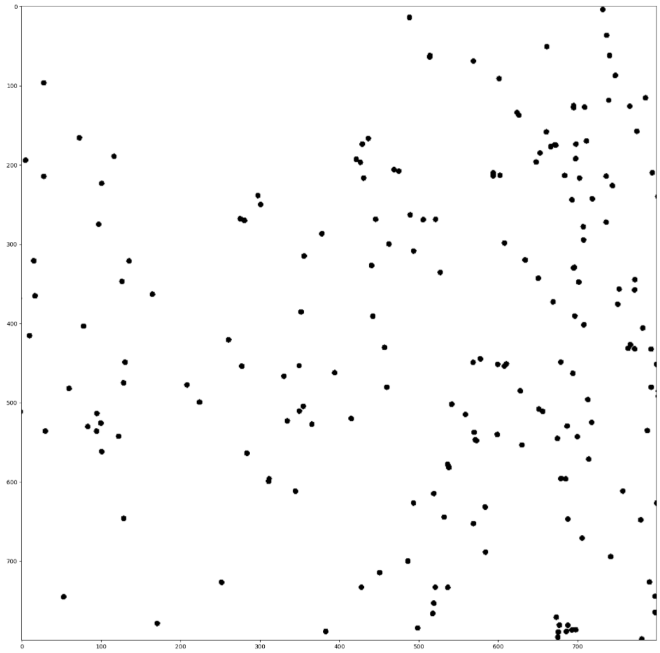


Figure 4 Sample Data Scatter

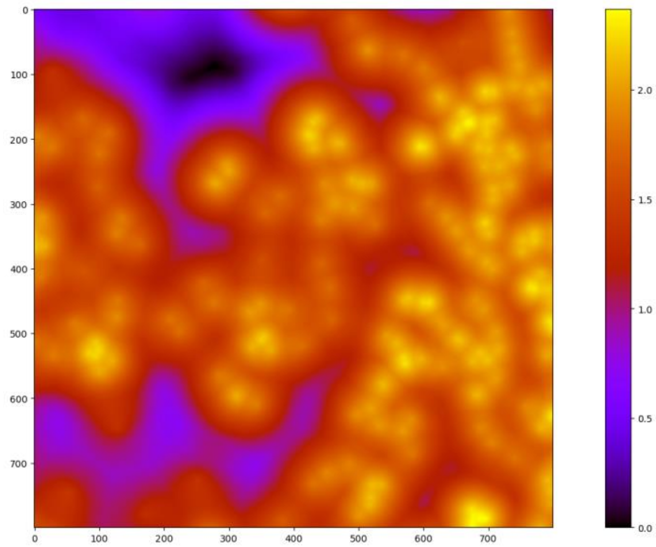
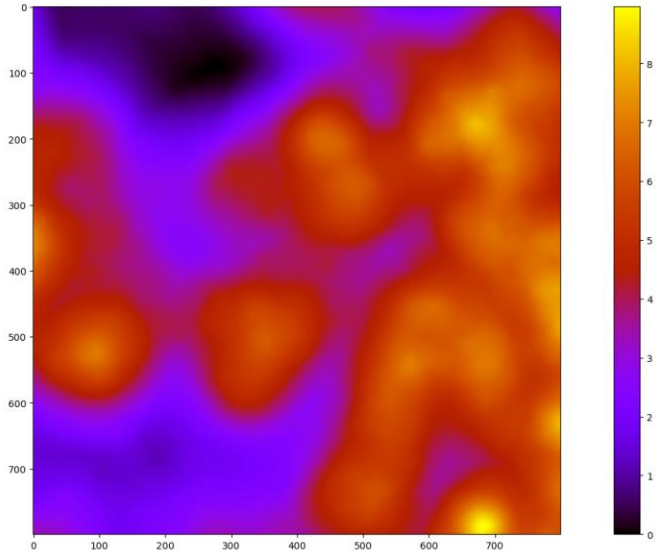


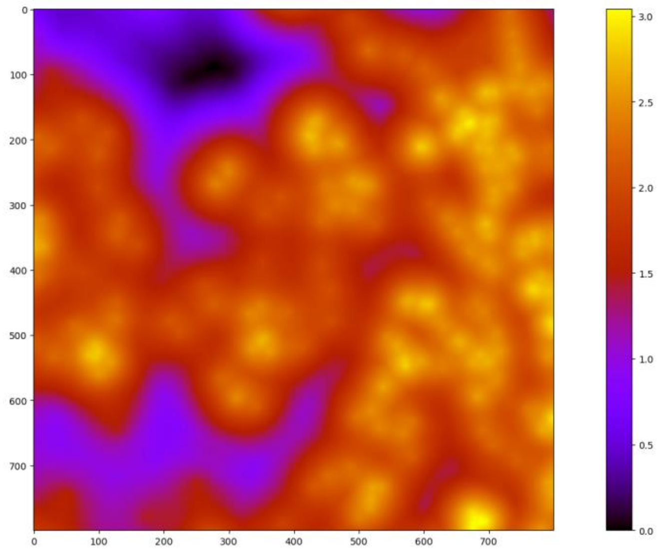
Figure 5 Sample Data Results

From the sample dataset shown in Figure 4, with the sample calculation with the parameters of exponent multiplier = 0.35, marginal multiplier = 0.8, search distance = 1500m, certain regions with multiple features could reach up to more than 1.0 (optimal single-feature ratio). These model dials will require further fine-tuning in order to close the distance to each application, with how much subsequent units are valued compared to the closer ones, how much value difference on each features due to their quality of services, and even their maximum distance.

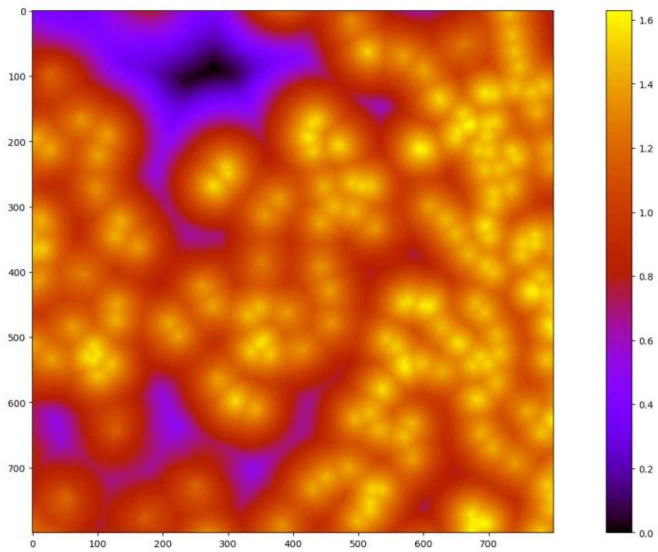
μ = 1.0



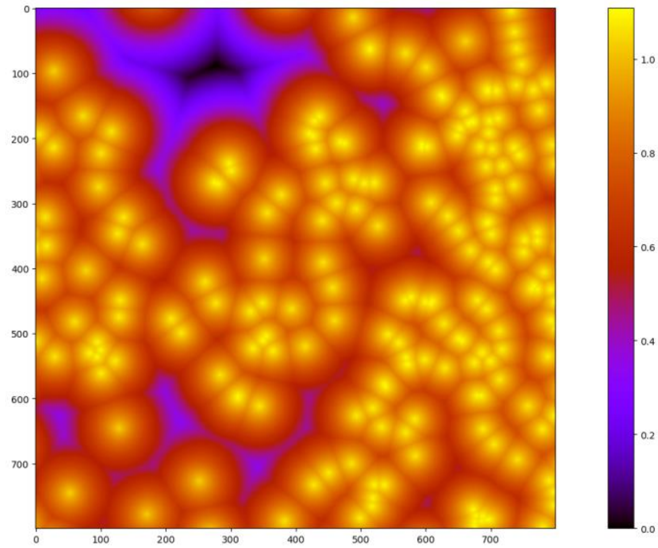
μ = 0.7



μ = 0.4



μ = 0.1



μ = 0.0

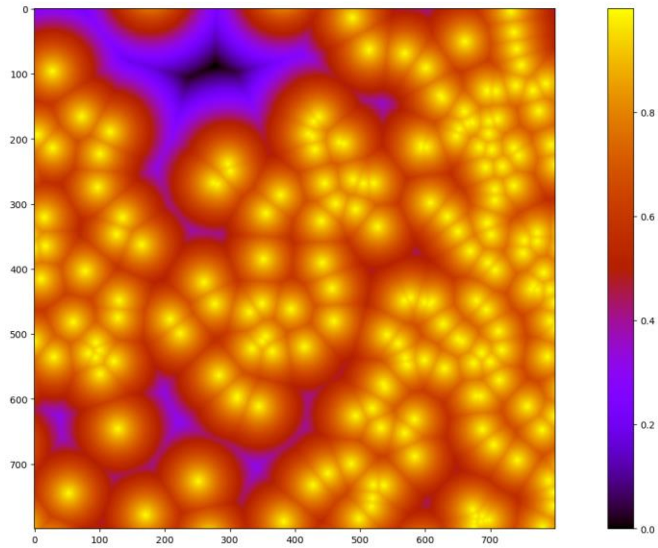


Figure 6 results of different marginal multiplier

Regarding the μ (marginal multiplier) in the calculations, as exhibited in Figure 6, in between the value of 1.0 (pure summation) and 0.0 (no summation, only closest) will provide different outputs and how the data is also read. As the higher constants will have higher points on locations that are scored in-between points, and smaller constants have a higher emphasis on closeness to any single point; Note that in this case all feature weight are the same at 1.0.

Some small illusory pattern that appears from the data visualization is that on low marginal multipliers, location in between two points where the distance is roughly the same seems to have a sudden decrease in value. However, upon investigation at the result's section/indices that is not the case, with the in-between values behaves as it should according to the laid-out process as shown on Figure 7

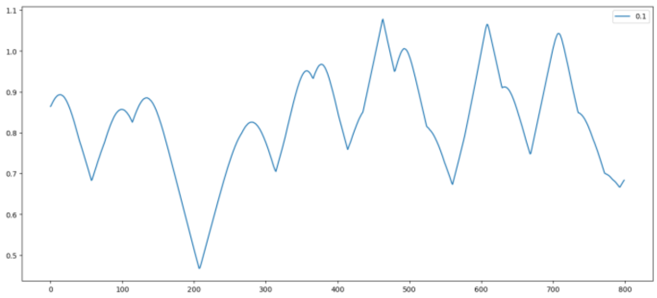


Figure 7 section index 300 of 0.1 u result

Accumulation of multiple features, their marginal multiplier, and their set individual weighting are detrimental to the result, with higher marginal values will also entail numerically higher scores, as well as different peak and valley locations as shown in Figure 8.

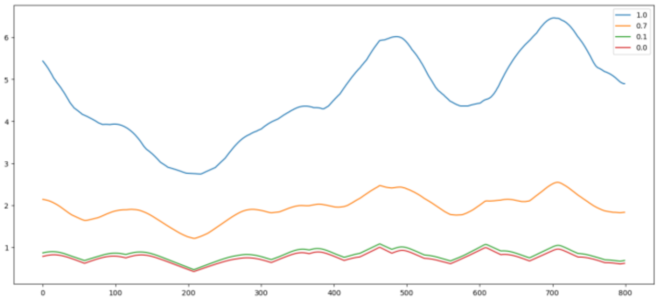


Figure 8 section index 300 of multiple u results

4. Notes and usage

This method is formed as a conjecture, and still bears the burden of proof on future study and application; As this is made by putting a common thought process into a quantitative framework.

References

Franke R (1982) Scattered data interpolation: tests of some methods. Math Comput 38:181-200

Shepard D (1968) A two-dimensional interpolation function for irregularly spaced data. In: Proceedings of the 1968 23rd ACM. ACM Press, New York

Miller W (2006) Delta Functions exercise solutions. Berkeley, California

Edgeworth F. Y. (1881) Mathematical Psychics, An essay on the application of mathematics to the moral sciences, London