ISSN: 2473-4901
v10 n6174

Duality in 2D Apportioning: A Site Suitability Case Study for Spatial Data Analytics

Peter Y Wu
wu@rmu.edu
Department of Computer Information Systems
Robert Morris University
Pittsburgh, PA 15108, USA

Abstract

A common application in spatial data analytics is site selection. In site suitability study, we often construct circular or ring buffers around the site location to analyze the vicinity. To determine its suitability, we assess the impact of the feature attributes around the site location by apportioning the parts of every feature included in the buffer. In our case study, we are commissioned to find a grocery store site with sufficient resident population in its vicinity. To facilitate for our search with test-and-verify strategy, the traditional approach will determine the proportion of population in the buffer around the potential site location. Instead, we exploit the duality of the census block polygon and aggregate the resident population of the census block to its centroid, turning the 2D polygons of census blocks into the points of their centroids. With the markers on the centroids symbolized by the resident population, the construct serves as a visual guide for our search. By aggregating the resident population to the centroid, the approach facilitates for our apportioning method to come up with an estimate of the resident population around the selected store site, verifying its suitability more efficiently.

Keywords: Site Suitability, Polygon Duality, Spatial Analytics, Geographic Information System, GIS.

1. INTRODUCTION

Maps are traditionally made for reference use. The geographic information system (GIS) has given rise to analytic mapping: maps are made for analysis to achieve various application purposes. Spatial data analytics is the study of large volumes of spatially related data. The modern GIS is currently trending toward the direction (Reddy, 2018; Dougherty et al., 2024).

Site suitability study is a common application in spatial data analytics. To determine whether or not a site location is suitable for a certain purpose, we gather the relevant information around the site. The data gathered is necessarily related spatially. We then construct a buffered region around the site to analyze the impact of relevant attribute information there. The buffered region is the site location extended with a circular buffer, or sometimes multiple layers of ring buffers for analysis of the impact with respect to proximity (Misra & Sharma, 2015).

Since the relevant attribute data may not be distributed evenly, the data may be provided in aggregation by partitioning regions, such as a polygon map layer over the area. We construct the buffered region around the site location, casting it over the polygon map to analyze the impact. Data apportionment is then the assessment of the portion of each polygon within the buffered region impacting the suitability (LaGro, 2013). While we present the relevant information to provide a visual guide for site selection, the data apportioning method will allow us to test and verify the site suitability. Repeating test and verify will lead our search to a feasible site location, but we need a more efficient apportioning method.

This paper introduces an apportioning method to exploit 2D duality of polygons. We convert the polygon map into a point map of the polygon centroids. We then aggregate the attributes of the

 ISSN: 2473-4901
v10 n6174

polygon feature to the centroid. In the absence of a way to exhaustively search for a suitable site, we use the point map of centroids to visually guide our search, and we use the centroids to facilitate our apportioning assessment for the suitability determination.

The next section has a brief survey of literature discussing our understanding of the current state of the art. Section 3 presents our case study of site selection for a grocery store in the downtown area, requiring certain resident population in the vicinity. Sections 4 and 5 describe our data gathering and how we set up our constructs to perform the spatial analysis. Section 6 discusses the benefits and limitations of our method. Section 7 closes with the summary of the paper.

2. LITERATURE SURVEY

Before the GIS started map making using the computer, maps were mostly for reference only. It was often too costly to make maps for the purpose of analysis (Schmidt, 1983). The GIS has become a very efficient and cost-effective tool for analytic maps (Heyward et al., 2006). Such is the trend for the GIS in spatial data analytics (Kanade, 2022). Site suitability is the assessment of factors and qualities to determine the site location for a particular activity. Site suitability study is a common application in spatial data analytics, facilitated by the use of the GIS (Jain & Venkata, 2007). Depending on the nature of the site selection project, there are varieties of different methods.

Spatial Environmental Model

The method usually applies to an area. The area is mapped and similar regions are identified with additional maps made. The maps incorporate factors relevant to study and a model is built for the purpose of comparison (Hopkins, 1977). The method requires comprehensive knowledge of the regions included and becomes onerous. It is not used often now.

Values Suitability Analysis

Geographical consideration in site suitability often extends to many other aspects. The ecological condition surrounding the site can be a natural extension, but aesthetic preferences as well as other human values, though subjective, also need to be taken into account quantifying costs and benefits (Reed & Brown, 2007).

Ordinal Priority Approach

The area is mapped according to the quality of the land, such as slope, soil type, vegetation, climate. Each quality is then quantified by a rating, and a method assigns the weight of each quality to determine the suitability (Mendoza et al., 2003; Melczewski, 2006).

In our site suitability case study, we have only one simple criterion: resident population. The issue involved then is the method of apportioning population in the neighborhood surroundings.

Apportionment Methods

The study of apportionment methods in the past decades has focused mostly on the politics of parliamentary seats allocation based on population (Kohler & Zeh, 2012; Koslap & Wilson, 2021) but not on the spatial distribution of population, which is related to political districting. That on the one hand depends on the politics of constitutional re-districting regulations, but it may also be because of the difficulty of apportioning based on 2D polygon intersections. When we have to rely on a visual guide to search for a feasible site location, all the more we need an efficient way to assess suitability so that we can easily repeat the test and verify procedure. For our case study, we propose a method described in the following sections.

we may need to augment the GIS with plug-in tools, or scripting support. If search for a suitable site location is necessary, there is often no exhaustive method available and we have to resort to search by repeated test and verify for suitability. Then we will want to have some kind of visual guide in the cartographic presentation and an apportioning method to assess the suitability [Misra & Sharma 2015].

3. CASE STUDY: SITE SELECTION

In our case study, we are commissioned by the city government in a project to revitalize businesses in the downtown area. While providing tax incentive to a potential grocery store, the city government asks us to provide a feasible store site location that meets the requirement of more than 1500 resident population within walking distance. We proceed to take up the project. Assuming the average walking speed of 3 miles per hour, the 10 minutes round trip walk means a range of ¼ mile, or 1320 feet. To create the study area for potential feasible site locations, we have the downtown area extended with the ¼ mile buffer. That is the area for a feasible store location. The only factor we need to consider is the resident population within the buffered zone. Figure 1 in the following page illustrates the study area, buffered around downtown, with the census blocks over the entire area.

 ISSN: 2473-4901
v10 n6174

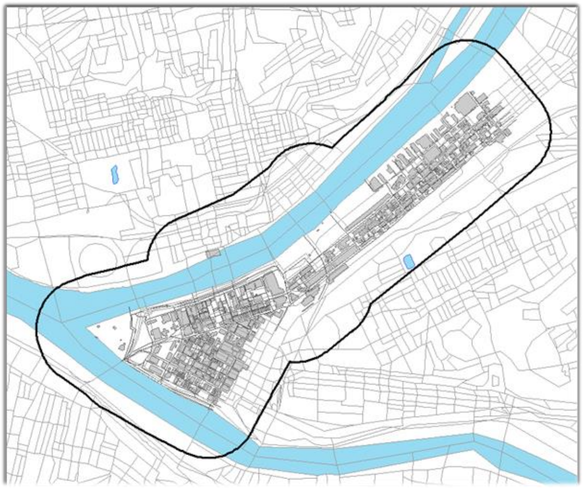


Figure 1: Study Area around Downtown

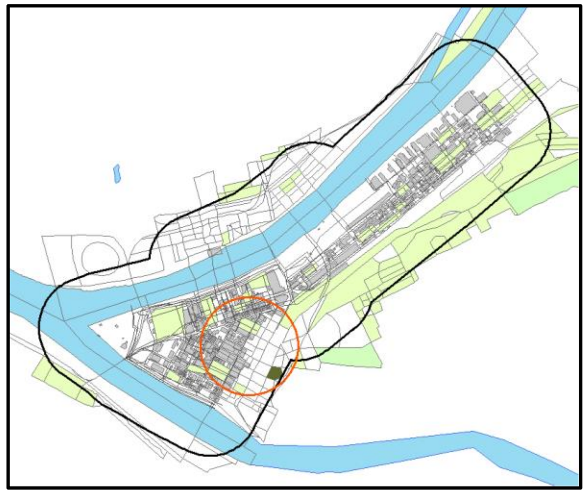


Figure 2. Census Blocks in a Choropleth Map.

4. DATA GATHERING AND VISUALIZATION

We gather the data of residential population from the U.S. census. The U.S. government holds census every 10 years. The resident population is counted in every census block. We use the study area to select the census blocks in and around the study area, with the information of resident population in every census block. The study area helps us to focus on the census blocks relevant to our project.

We can color shade every census block by the resident population to visualize the population pattern. The color shaded map is called a choropleth map (Dent et al., 2009). Since there is no full proof exhaustive way to iterate through all possible store locations, we want the choropleth map to be a visual guide for us in the search for a feasible store location. We may then test the location as feasible by verifying the residential population in the ¼ mile vicinity over 1500.

To estimate the resident population in the vicinity, we form a circular region of 1/4 mile radius at the store location. We need to prorate the apportionment of resident population in each census block that intersects with the circular region, assuming the population is distributed evenly within the census block and then we need to sum up all the portions of census blocks included. Figure 2 illustrates the study area with only the census blocks covering it, color shaded in the choropleth map by the resident population, and a ¼ mile radius circular graphic seeking a potential store location. The circular graphic requires us to calculate the prorated areas inside the circle to estimate the resident population.

5. SET UP FOR SPATIAL ANALYSIS

However, since we need to search for a potential site to test and verify, our approach becomes too cumbersome to be conveniently effective. We change our strategy to exploit the 2D duality of our census blocks map by aggregating the entire population of the census block polygon to the centroid of the polygon. The following describes our approach.

We convert the polygon map of census blocks to a point map of the centroids. Since we can easily calculate the (X,Y) coordinates of the centroid of each census block polygon based on its geometry, we can form the point map of the census block centroids. Each centroid is associated with the resident population of the entire census block. Instead of the choropleth map color shading the census blocks, we make a color marker at each centroid and color code the marker by the resident population of the census block. That provides a visual guide to our search for a feasible store location as good as the choropleth map. Figure 3 illustrates the point map of the centroids using markers to show the resident population of each census block, and we can move the circular graphic around in search of a feasible store site. Figure 3 is in the following page.

Now to calculate the apportionment of the resident population in the vicinity of any selected location, we will consider only the block centroids which fall entirely within the 1/4 mile radius circular buffer around the potential store location. If the centroid falls within the circular region, we include the entire population of the census block, and if the centroid does not fall inside, we disregard the census block. This approach is now

ISSN: 2473-4901
v10 n6174

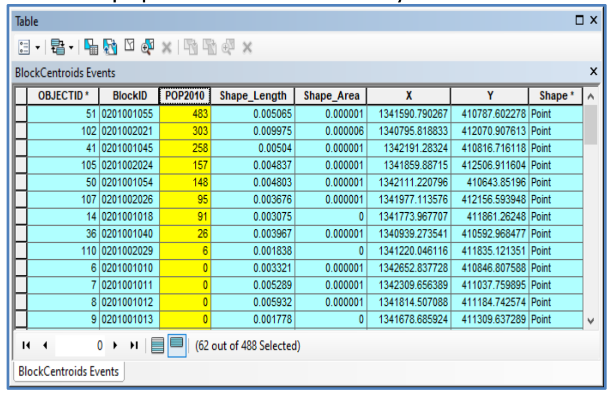
illustrated in Figure 4: the centroids within the circular buffer are high-lighted, noting the selection of census blocks to be included in the apportionment of resident population in the vicinity.

Figure 3. Census Block Centroids

Figure 4. Census Block Centroids Selected.

From the attribute data table for the Centroids point map layer, since the centroids inside the buffer circle are selected, we can easily sum up the Population now aggregated to the centroids. Figure 5 shows the attribute data table of the selected centroids with the population field high-lighted for summation.

Granted that the randomly distributed pattern of resident population within the census block, the approach still calculates a reasonable estimate of the population in the vicinity, but it can be effectively done with convenience. The strategy facilitates for our test and verify approach in the search for a feasible location, with an estimate of resident population in the vicinity.



Table

BlockCentroids Events

OBJECTID *	BlockID	POP2010	Shape_Length	Shape_Area	X	Y	Shape *
51	0201001055	483	0.005065	0.000001	1341590.790267	410787.602278	Point
102	0201002021	303	0.009975	0.000006	1340795.818833	412070.907613	Point
41	0201001045	258	0.00504	0.000001	1342191.28324	410816.716118	Point
105	0201002024	157	0.004837	0.000001	1341859.88715	412506.911604	Point
50	0201001054	148	0.004803	0.000001	1342111.220796	410643.85196	Point
107	0201002026	95	0.003676	0.000001	1341977.113576	412156.593948	Point
14	0201001018	91	0.003075	0	1341773.967707	411861.26248	Point
36	0201001040	26	0.003967	0.000001	1340939.273541	410592.968477	Point
110	0201002029	6	0.001838	0	1341220.046116	411835.121351	Point
6	0201001010	0	0.003321	0.000001	1342652.837728	410846.807588	Point
7	0201001011	0	0.005289	0.000001	1342309.656389	411037.759895	Point
8	0201001012	0	0.005932	0.000001	1341814.507088	411184.742574	Point
9	0201001013	0	0.001778	0	1341678.685924	411309.637289	Point

0 (62 out of 488 Selected)

BlockCentroids Events

Figure 5. Attribute Table with Centroids Selected

6. DISCUSSION

Our approach of aggregating the census block population to the centroid is different from prorating the population by the area of the census block inside the buffered region. However, we take into consideration only those census blocks that are within the buffered region. When the centroid is outside of the buffered region, the entire census block is not considered. We argue that the difference is acceptable for a reasonable estimate since the resident population is not necessarily evenly distributed within the census block, and is rather randomized.

The benefit of our approach exploiting the duality of 2D polygons is that we do not have to go through the cumbersome process of calculating the prorated portions of areas of each census block polygon within the buffered region. By aggregating the total population of the census block to the centroid, we can effectively come up with an estimate of the resident population in the vicinity with ease. That allows us to search for a store site location with ease, performing test and verify for feasibility strategy.

The strategy may also be generalized to apply to cases when the attribute information is randomly distributed and aggregated in polygons in the area. Our apportioning method turns the polygon map layer into a point map layer of centroids. By aggregating the attribute information to the centroids, we can assess the apportionment of the attribute information to the selected site location with ease. Further research effort may attempt to apply the methodology to other applicable situations.

We realize that there is a difference in the apportioning methods, but the difference is

ISSN: 2473-4901
v10 n6174

acceptable for the practice. and the ease of assessment supports our test-and-verify search strategy effectively. Another direction of further research effort should therefore be directed toward calibrating the method in comparison to other traditional approaches.

7. SUMMARY

In the site suitability problem of spatial data analytics, we have to assess the impact of certain attribute information distributed over the area in the vicinity around the site location. Noting that the attribute information is not evenly distributed, it is aggregated to the polygons partitioning the area. We can construct a buffered region to define the boundary of its vicinity and then determine the apportioning of the attribute by proration of the area of each polygon inside the buffered region. The approach is however cumbersome to handle and does not effectively support test-and-verify search. In our case study, we are asked to seek for a grocery store location with the requirement of certain resident population in the vicinity. The resident population information is aggregated in census blocks in the area. Instead of calculating the proration by the area of each census block in the buffered region, we convert the census block polygons into a point map of centroids, attributing the population of each census block to its centroid. We can then just consider those centroids that fall within the buffered region to come up with an estimate of the population in the vicinity. The difference in the assessment is acceptable, but the ease it allows offers us effective support for our test-and-verify search. The strategy can be generalized to similar cases of spatial data analytics.

8. REFERENCES

Dent, B.D., Torguson, J.S., Holder, T.W. (2009) Cartography: Thematic Map Design (6^{th} ed), McGraw-Hill.

Dougherty, P., Burden A. and Blitz M. (2024). Human by Design: How AI Unleashes The Next Level Of Human Potential. Technology Vision 2024, in Accenture-Tech-Vision. https://www.accenture.com/content/dam/accenture/final/accenture-com/document-2/Accenture-Tech-Vision-2024.pdf

Heywood, I., Cornelius S., Carver S. and Raju S. (2006). An Introduction to Geographical Information Systems, 2^{nd} edition, Pearson.

Jain, K. and Subbaiah Y.V., (2007). Site Suitability Analysis for Urban Development using GIS. *Journal of Applied Sciences* 7(18):2576-2583. https://scialert.net/abstract/?doi=jas.2007.2576.2583

Hopkins, L.D. (1977) Methods for Generating Land Suitability Maps: A Comparative Evaluation. *Journal of the American Institute of Planners* 13(4):386-400. https://doi.org/10.1080/01944367708977903

Kanade, V. (2022). What is Spatial Analysis? Definition, Working, and Examples. *Artificial Intelligence*, Spiceworks.com., August, 2022. https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-spatial-analysis/

Kohler, U. and Zeh J. (2012). Apportionment Methods. *The Stata Journal* 12(3):375-392. https://doi.org/10.1177/1536867X1201200303

Koslap, K. and Wilson S. (2021). How Apportionment is Calculated. Published in *Census Blogs*, April 26, 2021. https://www.census.gov/newsroom/blogs/random-samplings/2021/04/how-apportionment-is-calculated.html

LaGro, J.A. (2013) Site Analysis: Informing Context-Sensitive and Sustainable Site Planning and Design. John Wiley & Sons, Inc. Hoboken, NJ.

Melczewski, J. (2006) GIS-based Multicriteria Decision Analysis: A Survey of the Literature. *Journal of Geographical Information Science* 20(7):703-726. https://doi.org/10.1080/13658810600661508

Mendoza, G., Anderson, A. & Gertner, G. (2002) Integrating Multi-criteria Analysis and GIS for Land Condition Assessment: Evaluation and Restoration of Military Training. *Journal of Geographic Information & Decision Analysis* 6(1):1-16.

Misra, S.K. & Sharma S. (2015). Site Suitability Analysis for Urban Development: A Review. *International Journal on Recent and Innovation Trends in Computing and Communication* 3(6):3647-3651 (ISSN:2321-8169). https://ijritcc.org/index.php/ijritcc/article/view/4511

Reddy, G.P.O. (2018). Spatial Data Management, Analysis, and Modeling in GIS: Principles and Applications. In: Reddy, G., Singh, S. (eds) *Geospatial Technologies in Land Resources Mapping, Monitoring and Management*, vol.21. Springer, Cham. https://doi.org/10.1007/978-3-319-78711-4_7

ISSN: 2473-4901
v10 n6174

Reed, P. and Brown, G. (2007) Values Suitability Analysis: A Methodology for Identifying and Integrating Public Perceptions of Ecosystem Values in Forest Planning. *Journal of Environmental Planning and Management* 46(5):643-658. https://doi.org/10.1080/0964056032000138418

Schmidt, A.H. (1983). Automated Cartography. *Computer Graphics*, T.L. Tunii (eds), Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-85962-5_30