Automatic Generation of Station Catchment Areas:

A comparison of Euclidean distance transform algorithm and location-allocation methods

Ting(Grace) Lin, Richard Palmer, Jianhong (Cecilia) Xia and David A. McMeekin
Department of Spatial Sciences
Curtin University
Perth, Australia

Abstract—A train station catchment area can be generated in two ways: by surveying train users or by modelling methods. The former method is time consuming and labor intensive. In this paper attention was given to develop methods that automatically generate station catchment areas. This study's aim was to compare two modelling methods: the Euclidean distance transform / Voronoi diagram generation method and a location-allocation method for automatically generating station catchment areas. A case study of the Perth Metropolitan area, in Western Australia, was used to implement these two methods. The results from these two methods are consistent and the methods demonstrate robustness for understanding the nature of station catchment areas and provide useful insights for public transport planning in Perth.

Keywords-station catchment area; Euclidean distance trnasform; Voronoi diagram; Location-allocation

I. INTRODUCTION

Train station catchment areas can be defined as an area in which local residents come from, whom could potentially use the train services. Many studies have been conducted in measuring station catchment areas[1, 2]. However, the vast majority of methods were developed based on the traveler's origin information, collected using survey methods that are time consuming and labor intensive. Therefore, more efficient methods, such as automatically generating the catchment area are needed. The aim of this study is to compare the two methods: the Euclidean distance transform and the location-allocation method in automatically generating catchment areas. The analysis of location-allocation methods are based on the network distance and potential demand, such as population in each district, while the Euclidean distance transform algorithm is based on the Euclidean distance and patronage of stations. A month's SmartRider (the Western Australian Public Transport system's electronic ticketing system) train service usage data were used to calculate the train services' demand using a circle generation method developed in this study. The results of this research may be used to understand the latent demand of train stations and the impact on train station services.

II. BACKGROUND

Catchment area studies, conducted from transportation perspectives, can be summarized into three areas:

A. Catchment area measure

Train station catchment areas were generally measured using two different categorizing methods: 1) Deriving a catchment area based on geocoded trip survey data and calculating a boundary by convex hull methods [2]. Generally, only 90% of the trip data were used within the study in order to remove the outliers [1, 3]. 2) Deriving a catchment area using network analysis techniques [4]. A catchment area can be a boundary covering a fixed walking or driving distance or travel time based on path networks [5]. In addition, a catchment area can be generated based on optimizing the utility of stations. This method's base assumption is that the further people have to travel to reach a station, the lower probability they will use it. The method was designed to maximize a facility, such as a train station's patronage [6, 7]. 3) Deriving a catchment area by simply placing a buffer circle around a station. For example, Studies by [1] and [5] used 400, 800 and 1200 meter buffers to define walking distance to public transport.

B. Catchment area characteristic analysis

The size and shape of catchment areas were determined by the characteristics of travelers, their trips and train stations. For example, Sanko and Shoji (2009) found that the station's catchment area size can be influenced by travel modes, train frequency and station intervals [8]. Stations with express train services tend to have larger catchment areas than stations without this service [8]. A walking catchment area is more

This research was supported by an ARC Linkage grant (LP110201150) and sponsored by Australian Research Council, Department of The Planning, Department of Transport and The Public Transport Authority, Western Australia.

likely to be concentric and smaller than other travel mode catchment areas [1]. The station's location may also affect a catchment area size. A terminal station usually has a larger catchment area than a station along a train line [9]. Willingness-to-walk may also differ by age, gender, land use, safety, weather, and the cost and availability of parking [10].

C. Applications of catchment area

Catchment areas have been extensively used to understand travel behavior and the impact of transit services on this behavior. For example, research has been conducted to understand market share and latent demand of transit services based on station catchment area analysis [11]. The catchment area can be used to understand competition between railway stations [12]. The overlaps between station catchment areas may indicate a train station oversupply [12]. Attention was also drawn to understand the relationship between proximity to a station or its catchment areas and property value. The closeness to a train station can be considered a positive factor affecting property prices [13].

This paper focuses on the catchment area measure, especially the methods that automatically generate station catchment areas.

III. Methodology

A. Circle generation methods

In this paper, the station attractiveness was determined by station patronage and the spacing of stations. This was represented circularly. The radius of the circle was calculated as:

$$R_i = \frac{P_i}{C} \tag{1}$$

$$C = Max\left\{\frac{P_i + P_j}{D_{ij}}\right\} \tag{2}$$

where:
R_i is the radius of the circle of station *i*;
C is the non-zero constant.
P_i is the patronage of the station *i*;
P_j is the patronage of the station *j*;
Station *i* and *j are adjacent stations;* and
D_{ij} is the Euclidian distance between Station *i* and *j*

The station attractiveness is directly proportional to the station patronage. Station patronage was calculated from SmartRider boarding data for weekdays in September 2011. The data were provided by Department of Public Transport Authority, Western Australia.

The constant *C* was used to rescale the patronage data in order to avoid overlapping between circles. It is the largest ratio between the number boarding at station *i* and *j* and the Euclidian distance between station *i* and *j.*

B. Euclidean Distance Transform / Voronoi Diagram Generation algorithm

The distance transform *d(p)* is a mapping of a binary image *I(p)* ∈ {1,0} comprised of foreground (or object) points and background points, to an image where every foreground point *p* is given the distance to its closest background point *q*. Formally,

$$d(p) = \begin{cases} 0 & I(p) = 0 \\ \text{argmin}_{I(q)=0} \| q - p \| & I(p) = 1 \end{cases} \tag{3}$$

where $\|q - p\|$ is some distance metric between *q* and *p*.

Though many different metrics can be used to compute a distance transform (DT), the Euclidean distance (the L_2 norm) is preferred for many applications because it provides the exact isotropic (radially symmetric) real world geometric distance between two points [14-17].

$$\| q - p \|_2 = \sqrt{(q_z - p_z)^2 + (q_y - p_y)^2} \tag{4}$$

Though more complex to compute, the advantage of the Euclidean distance metric over other kinds of distance metrics (such as the City Block or Chessboard distance) is that it gives the true shortest distance between points. The use of the Euclidean metric also allows for the algorithm to be easily modified to generate the Voronoi or Dirichlet tessellation where the image is partitioned into discrete regions and each region is mapped to its closest foreground point.

When generating the Voronoi diagram for the train stations' residential catchment areas, the effect of varying the radii of the point circles representing the train stations is to influence the size of the surrounding region corresponding to the catchment area of the station. This is because distances are calculated to the nearest points on the boundary of the closest foreground points.

Generation of the Voronoi diagram proceeds in two phases. In the first phase, 1D squared Euclidean distances are computed along each row of the image (requiring a maximum of two passes per row). After this phase, each point corresponding to a background pixel contains the squared Euclidean distance to its closest foreground point in the same row. In the second phase, squared Euclidean distances are computed along each column of the image. Each column is parsed in both directions to compute the influencing regions of each laterally offset foreground pixel. The resulting list of regions in each column gives a mapping to the closest foreground points, and then to the value used to uniquely label the foreground point associated with a train station. The list of regions is then parsed to set this unique value to each of the mapped background points. As shown in Fig. 2 the output of this process is to generate an image with each pixel having the same value as used to color the foreground circular points in the input image.

C. location-allocation methods

Location-allocation is "the simultaneous location of central facilities and the allocation of dispersed demand to them, so as to optimize some objective function" [18]. For example, giving N candidate facilities and M demand points with a weight, such as population, the location-allocation methods can maximize the attendance at facilities from the demand points [19].

P-median is one of the earliest formulated standard location-allocation problems [18]. The key idea of p-median is to minimize the total shortest demand weighted cost between the facilities and demand points.

However, p-median methods on a general network are "hard" problems (NP-hard), which means that in order to seek the best combination of demand points and allocate them to P facilities, enormous computation is required [21, 22]. Therefore, a number of heuristic approaches have been developed to solve this problem. Nagy and Salhi [23] classified them by the solution methods employed to three groups. The first one is clustering-based sequential methods. This method partitions the demands into clusters based on the weighted Euclidean distance, and then the Vehicle route problem (VRP) is further applied to assign a route for each cluster. The second method is iterative and iterates between sub-problems or phases, such as the global phase and regional phase (Global Regional Interchange Algorithm (GRIA)) and feeds information between the two phases. The third method is a hierarchical method. Location was considered as a main problem and routing as a sub-problem. The main algorithm focuses on solving the location problem and then solving the routing problem as a subroutine in each step. Nagy and Salhi [23] conceived that hierarchical algorithms may be a better solution to a real location-allocation situation.

The methodology applied in the paper is p-median whose equation can be defined as[24]:

$$Min(\sum_{i\in M}\sum_{j\in N} W_i d_{ij} y_{ij}) \qquad (5)$$

where:
N is the number of facilities;
M is the number of demands;
y_{ij} is the decision variables, y_{ij} =1 if demand i is located in facility j;
d_{ij} is the cost (network distance) between demand and facility; and
W_i is the weight in the demand such as the number of the population.

The heuristics method (GRIA) was applied to solve the p-median problem in this paper. The ArcGIS software was used to implement the method.

IV. Methods of Implementation

The above developed methods in this study were implemented using a case study of the Perth Metropolitan area in Western Australia. There are five main train lines (Armadale, Fremantle, Joondalup, Midland and Mandurah) and one spur line (Thornlie) and 69 train stations on 168 kilometers of track (See Fig. 1) [25].Circles, representing the attractiveness of stations were generated for each station using equation 1 and 2 and ArcGIS software. The circles created were in a colored vector format. However, in order to be used as an input file for the catchment area calculation using the Euclidean distance transform algorithm, the colored vector map was converted into a Grayscale raster image (a PNG file) with a range of 0-255 using ArcGIS. Based on the circle image, the Euclidean Distance Transform / Voronoi generation algorithm was implemented in C++ to generate the catchment area for each station automatically according to the general algorithm described above. The location-allocation method (Maximize attendance) was implemented using ArcGIS. Instead of giving a boundary of a catchment area, the catchment area was illustrated using a supply and demand matrix or origin and destination matrix: a straight line between a station and a demand point, in our case is the centroid of statistical area 1 (SA1). Although it was illustrated as straight line, the distance

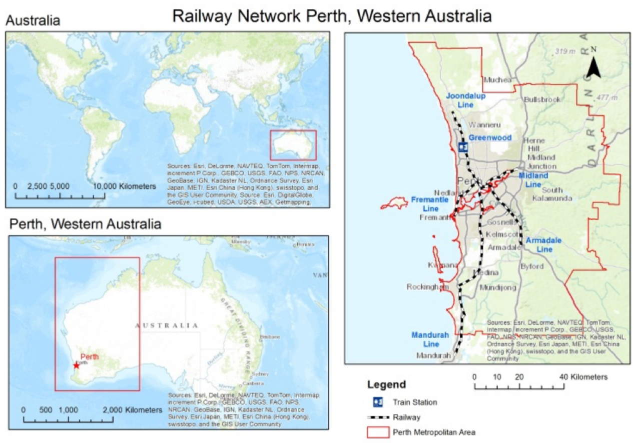

Figure 1. Study area

between the station and the centroid was calculated as the network distance.

V. Results

Fig. 2 illustrates the results calculated using the three methods. The circles were created in Greyscale from circle generation methods. However, for visualization purposes, the circles were highlighted in red with a white outline (See Fig. 2). The grey catchment area of each station in the map's background was generated using the Euclidian distance transform algorithm. The Greyscale color is consistent with the color of the circle. The colored supply and demand matrix line was drawn using location-allocation methods. The demand points show the latent demand areas. Travelers from these areas could potentially use the train services. As can be seen in Fig. 2, the catchment area size is not directly proportional to the attractiveness or patronage of the station. For example, the Murdoch station along the Mandurah line is the most attractive station with the highest boarding numbers. However, its catchment area is relatively small. In addition, the catchment areas calculated from two different methods are mostly consistent. Small differences between them are generally located at the areas along the borders of station catchment areas (See Fig. 2).

VI. DISCUSSIONS AND CONCLUSIONS

The two catchment generation methods are principally similar and both are based on Delaunay triangulation. The differences between the two methods are 1) distance calculation methods where the Euclidian distance transformation methods used Euclidian distance while location -allocation methods were based on network or Manhattan distance. 2) Factors considered for optimizing the catchment areas. For the Euclidian distance transformation methods, the number of stations available, the patronage of the station represented by the size of the circle and the distance between stations were taken into consideration. In this study, the distance between the stations was calculated from the outline of the circle instead of the circle's center. Therefore, the size of circle can be taken into account when determining the catchment area size. The size of the circle was deliberately increased twice in order to test the algorithm. The catchment area was slightly increased especially for smaller catchment areas (See Fig. 3). This may mean that the algorithm is more sensitive to the size of the circle in generating a small catchment area. For the location-allocation method, the number of stations available, population in each SA1, the distance between stations and distance between stations and demand points were considered. In summary, the main difference between these two methods in terms of factors is that the Euclidean distance considered the real patronage of stations, while the location-allocation method considered the population in SA1 that is latent demand.

From these results, it has been identified that the size of the catchment area is not directly proportional to the attractiveness or patronage of the station. Therefore, it can be concluded that the catchment area size was influenced by the station attractiveness. However, other factors, such as station location, travel modes, and train frequency discussed in the Background section, may also affect the station catchment area size. In addition, the station catchment areas are more likely to overlap with each other. For example, in Fig. 3, the point data shows the geocoded origin location of Park and Ride (PnR) users of Midland, Murdoch and Meltham station in 2008. The Murdoch station PnR users were from a much broader area. In order to capture the nature of the catchment area, more complex algorithms need to be developed in the future, which can consider more factors and allow the overlapping between catchment areas.

In summary, this study compared the two automatic catchment area generation methods: the Euclidian distance formation algorithm and the location-allocation methods. The results from these two methods were consistent and the methods were shown to be robust for understanding the nature of station catchment areas and will provide useful insight for public transportation planning in Perth.

ACKNOWLEDGMENT

The authors would like to thank Dr Todd Robinson for discussing the ideas of the paper. We are very graceful to PTA (The Public Transport Authority) and DoP (The Department of Planning) for their ongoing support for this study. The views expressed in this article are those of the authors, and do not necessarily reflect the policies of any organization

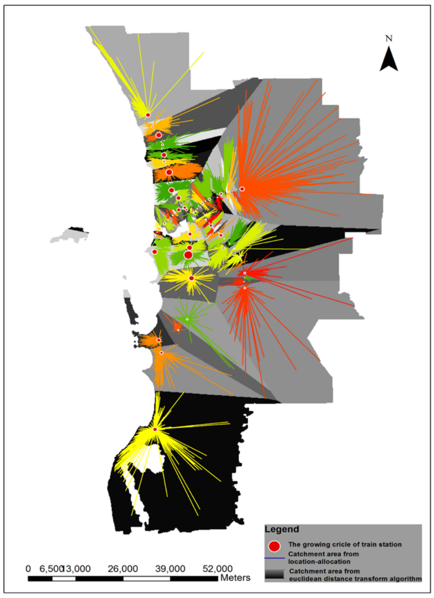

Figure 2. Catchment areas of Perth train stations

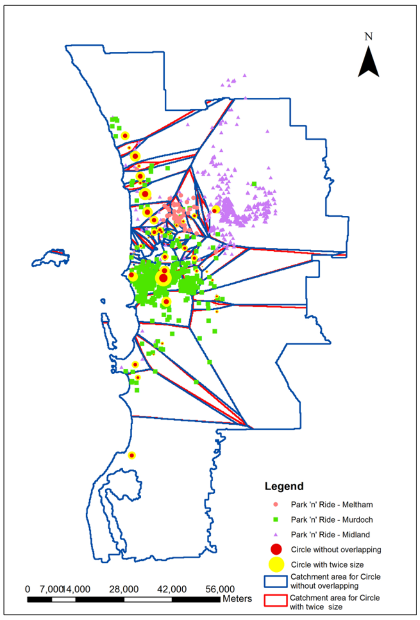

Figure 3. The comparison of catchment areas based on two different sizes of circles using the Euclidean distance transform algorith

REFERENCES

[1] R. Cervero, A. Round, T. Goldman, and K. Wu, *Rail access modes and catchment area for the BART System*, Transportation Centre, The University of California, USA, 1995.

[2] P. Durr, K. Graham, and S. Eady, *GIS mapping of cattle market service areas using the National Livestock Identication System*, Australian Animal Health Laboratory, Geelong, Australia,CSIRO Livestock Industries, Armidale, Australia, 2010.

[3] Irvine. "Analyzing the market share of commuter rail stations using LEHD Data," 10/3/1014, 2014; http://onlinepubs.trb.org/onlinepubs/Conferences/2011/Census/Presentations/10-26_830-10amCervenka/2Eichler.pdf.

[4] A. Landex, and S. Hansen, "GIS-based approaches to catchment area analyses of mass transit," in ESRI International User Conference, 2009.

[5] A. Etman, C. B. Kamphuis, R. G. Prins, A. Burdorf, F. H. Pierik, and F. J. van Lenthe, "Characteristics of residential areas and transportational walking among frail and non-frail Dutch elderly: does the size of the area matter?," *Int J Health Geogr,* vol. 13, pp. 7, 2014.

[6] M. J. Hodgson, "A location—allocation model maximizing consumers' welfare," *Regional Studies,* vol. 15, no. 6, pp. 493-506, 1981/12/01, 1981.

[7] E. Hillsman, "The p-median structure as a unified linear model for location-allocation analysis," *Environment and Planning A,* vol. 16, pp. 305-18, 1984.

[8] N. Sanko, and K. Shoji, "Analysis on the structural characterisitics of the station catchment area in Japan," in 11th Conference on Competition and Ownership in Land Passenger Transport, Delft University of Technology, The Netherlands, 2009.

[9] M. Duncan, and R. K. Christensen, "An analysis of park-and-ride provision at light rail stations across the US," *Transport Policy,* vol. 25, pp. 148-157, 2013.

[10] E. Guerra, R. Cervero, and D. Tischler, "The half-mile circle: does it represent transit station catchments?," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2276, pp. 101-109, 2012.

[11] A. Landex, and S. Hansen, "Examining the potential travellers in catchment areas for public transport.", 26th Annual ESRI International Conference.

[12] M. Givoni, and P. Rietveld, "Access to rail in urban areas–examination of the number of stations", School of Geography and the Environment,Oxford, 2010.

[13] G. Debrezion, E. Pels, and P. Rietveld, "The impact of railway stations on residential and commercial property value: a meta-analysis," *The Journal of Real Estate Finance and Economics,* vol. 35, no. 2, pp. 161-180, 2007.

[14] C. R. Maurer Jr, R. Qi, and V. Raghavan, "A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 25, no. 2, pp. 265-270, 2003.

[15] D. G. Bailey, "An efficient euclidean distance transform," *Combinatorial Image Analysis*, pp. 394-408: Springer, 2005.

[16] H. Breu, J. Gil, D. Kirkpatrick, and M. Werman, "Linear time Euclidean distance transform algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 17, no. 5, pp. 529-533, 1995.

[17] R. Fabbri, L. D. F. Costa, J. C. Torelli, and O. M. Bruno, "2D Euclidean distance transform algorithms: A comparative survey," *ACM Computing Surveys (CSUR),* vol. 40, no. 1, pp. 2, 2008.

[18] M. F. Goodchild, "ILACS: A location-allocation model for retail site selection," *Journal of Retailing,* vol. 60, pp. 84-100, 1984.

[19] Esri. "Algorithms used by network analyst," May 8, 2014; http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Algorithms_used_by_Network_Analyst/004700000053000000/.

[20] D. o. C. S. a. I. E. NTNU. "Location allocation problem," May 8, 2014; http://www.csie.ntnu.edu.tw/~u91029/LocationAllocationProblem.html.

[21] T. Kikuno, N. Yoshida, and Y. Kakuda, "NP-completeness of some type of p-center problem," *Discrete Applied Mathematics,* vol. 2, no. 4, pp. 361-363, 12//, 1980.

[22] N. Mladenović, J. Brimberg, P. Hansen, and J. A. Moreno-Pérez, "The p-median problem: a survey of metaheuristic approaches," *European Journal of Operational Research,* vol. 179, no. 3, pp. 927-939, 6/16/, 2007.

[23] G. Nagy, and S. Salhi, "Location-routing: issues, models and methods," *European Journal of Operational Research,* vol. 177, no. 2, pp. 649-672, 3/1/, 2007.

[24] G. Sierksma, *Networks in action : text and computer exercises in network optimization*, MA: Boston, MA : Springer US : Imprint: Springer, 2010.

[25] Department of Infrastructure and Transport, *Understanding australia's urban railways*, 2012; http://www.bitre.gov.au/publications/2012/files/report_131.pdf.