Commercial Activity Cluster Recognition with Modified DBSCAN Algorithm: A Case Study of Milan

1st Jiabin Wei
College of Architecture and Urban Planning
Tongji University
Shanghai, China
wjbkuge@yahoo.com

2nd Shiyu sun
Google Cloud Platform
Google Inc.
Seattle, USA
sshiyu@google.com

***Abstract*— The clusters of stores and shops in the city are the main spatial carrier for commercial activities. For urban planners, a deep and clear understanding of the present aggregating features and the commercial activities is the fundamental premise for formulating a rational and promising planning. Nowadays, the volunteered geographic information, like POIs, provides researchers a more complete and realistic data source to analyse the commercial agglomerations. Yet, few of the researches pay attention to the scale of the commercial agglomerations while the majority of researches use density estimation method to visualize and describe the commercial agglomerations of different activity types at same scale. This paper aims to propose a modified DBSCAN method to analyse the distribution structures of commercial activity clusters through multiple scales, so as to find the optimum parameters ε and minPts to identify the unique aggregating features for each type of activity. The proposed DBSCAN is able to determine the global minimum points (minPts) automatically by detecting the "elbow" of the maximum cluster groups change curve through a series combination of ε and minPts. With the global optimum minPts, this modified DBSCAN will further find optimum ε from where the commercial activities form stable aggregations. In this paper, the commercial activities in Milan is taken as an example. Overall, 149234 POIs from the Milan Bureau of Industry and Commerce and Google place service are collected and be further classified into 25 categories. The result of the analysis shows that 1) commercial activities show five different types of spatial patterns: central aggregation pattern, ring around center pattern, high-density aggregation distribution, disperse distribution pattern and hierarchical distribution pattern. 2) Bars and clothing stores have the highest aggregating density of 2.7 POIs per hectare, while takeaway and repair activities have the lowest density. 3) Beauty stores and health service have the smallest unit cluster size around 3ha, the supermarkets and fuel stations have largest unit cluster size. 4) the spatial shapes of different activity agglomeration areas are varied.**

Keywords—modified DBSCAN, global optimum minPts, cluster recognition, Milan, commercial activity

I. Introduction

In all kinds of society, individuals always want to aggregate together in order to get benefits from the advantages of the cluster, but there are always resistance that dispersing individuals away. Cluster is a phenomenon of centralized distribution of related enterprises, institutions and organizations in a certain field [1]. As one of the most important urban economic activity categories, commercial activities can be considered as a special case of industrial agglomeration and become more and more important in the 21st century [2]. Economists have developed essential theories on the formation of enterprises and manufacturing industries agglomerations. However, from the perspective of urban planning and land management, apart from the causes of agglomeration, the distribution patterns and spatial structure of commercial activity agglomeration are also worth studying. These spatial findings play a guiding role in the decision-making process on urban planning, such as the location selection of a new industrial park, the commercial structure of a shopping street, the renewal of an old town and so on. Thus, a scientific understanding of commercial agglomerations will give a full play to the driving role of commercial activities in urban life and urban economy.

The development of information technology has brought new data source and research method to urban studies. In the past, researches were usually based on small sample of data observed from field work and investigation. Today, volunteered geographic information (VGI), such as google, Bing, OSM, provides a more comprehensive and massive data of economic activities[3]. In addition, more efficient and faster data processing capabilities make it possible to collate, describe and analyse big data with complex algorithms. Many researchers have tried to use data science algorithms to visualize and quantify the urban commercial activities, so as to describe the distribution of commercial activities more accurately and extract the characteristics and rules behind the distribution. In the study of commercial spatial distribution, different types of commercial activity tend to present different aggregation states at different scales. How to effectively identify the aggregation scale of a specific commercial activity type is the key factor to the study of commercial spatial distribution. This paper aims to explore aggregating characteristics of urban commercial activities with a modified DBSCAN algorithm, which can be used to automatically select global optimum minimum point parameter and carry out continuous clustering analysis in a certain ε range to find several local optimum ε. Through multiple observation scales, a more accurate recognition on the commercial activity agglomerations for each commercial activity type can be derived.

II. Related Work

As a typical representative of VGI data, POI can directly reflect the type and distribution of urban commercial

978-1-7281-0846-9/19/$31.00 ©2019 IEEE

activities, and it is also easier to obtain than other data. Therefore, POI data is commonly used for spatial analysis of urban economic activities at the fine scale level [4]. At present, researches on aggregation distribution of urban commercial activities with POI mainly rely on spatial statistical methods for raster data and sample data, which are commonly used in the GIS, such as Moran index, Getis index and so on. For example, Chen, Liu, and Liang [5] and Yu and Ai [6] firstly rasterized and visualized POI data with Kernel Density Estimation, then further analyse the spatial aggregation of commercial activities with Getis' index, local Moran index. However, the recognition result is directly affected by the partition scale of statistical cells or grids [7]. Thus, the process of dividing and re-sampling point data into grid or raster data might lead to some inevitable errors. More importantly, commercial activities, like ecological activities, have the characteristics of scale dependence [8], so they tend to show different aggregation characteristics at different scales. However, as the rasterization or re-sampling method can only be used to analyse the distribution of business activities at one scale, they might omit some import characteristics in other scales.

The cluster can be defined as the area with higher density than the surroundings. To identify the cluster of a point set, density-based algorithms are usually considered as the best choices. There are already some effective density-based algorithms. One is WaveCluster algorithm. Sheikholeslami, Chatterjee, and Zhang [9] applied wavelet transformation to cluster points and it is capable of detecting clusters at different grid scales. Another popular density-based clustering algorithm is SNN proposed by Ertoz, Steinbach, and Kumar[10], which measures the density and similarity by the shared nearest neighbors number of each points. However, WaveCluster uses grid to caculate the density, which might encounter the same accuracy problem as rasterization process. As for SNN, it is not very convinent to use because three parametrs need to be set beforehand.

This paper introduces DBSCAN clustering algorithm to recognize the aggregation area of commercial activities in order to avoid the rasterization of the point data and to detect clusters in multiple scales. The DBSCAN algorithm is a density-based clustering algorithm, mainly applied to the clustering and pattern recognition problem of large spatial databases [11]. Nowdays, it is widely used in object recognition, gene expression analysis, star detection research filed [12-14] . However, There are some limitations on the basic DBSCAN algorithm. For example, it needs to set parameter searching radius (ε) and parameter minimum points (minPts) manually. In addition, the algorithm is difficult to identify and distinguish clusters with different densities [15]. Researchers often make some modifications and improvements to DBSCAN according to the specific problems. This paper aims to find the optimum observation scale of a set of commercial activity points. Therefore, an improved DBSCAN algorithm is needed to be capable to detect clusters under different searching radius or scales and smaller clusters in large clusters. Pavlis, Dolega and Singleton [16] tried to find the retail centers in different cities by decomposing the retail points into subgraphs and iteratively implementing DSBCAN on the merged areas with similar densities. Liu, Zhou, and Wu [17] and Naik Gaonkar, and Sawant [18] proposed to select several possible ε based on the distribution characteristics of point set from k-dist plot, and then execute the DBSCAN with possible ε so as to detect different density aggregation in the point set. Liu, Huang and Gao [19] combines the advantages of graph decomposition and k-dist plot. The proposed algorithm could generate optimum eps and minPts automatically for each activity zones based on k-dist. However, it failed to detect the heterogeneity inside the clusters. Although all the mentioned modified DBSCAN algorithms above provide a scientific way of finding optimum ε, there is still a lack of reasonable methods to define the global or local parameter minPts. We see potential possibilities to detect clusters of different densities at multiple scales with global optimum minPts.

In this paper, a new method is proposed to determine the global optimum minPts of DBSCAN automatically and, with global optimum minPts, the local optimum ε. This modified DBSCAN can not only aid researchers to quickly find the appropriate minPts and ε of a complex point set, but also enable the detection of clusters at different scales for a complex data set. The clustering scales could reveal the spatial distribution laws of commercial activities in the city. The commercial activities in Milan are taken as an example to test the proposed modified DBSCAN algorithm.

.

III. Modified Dbscan Algorithm

In this part, basic DBSCAN and the main logic of finding the optimum clustering result with a multiple scale clustering will be introduced. The key prerequisite of executing multiple scales clustering is to fix a minPts. Then, a modified DBSCAN is proposed to determine the global optimum minPts by maximum cluster number change 'elbow'. Finally, local optimum ε can be extracted based on the shape of the cluster number change curve and the clusters of any point set can be clearly identified under global minimum minPts and local optimum ε.

A. *Introduction for DBSCAN*

DBSCAN(density-based spatial clustering of applications with noise) is a density-based data clustering algorithm. Given a set of points in some space, it divides points into three types: core points, reachable points and outliers and use their relationship to make clustering. To generate the clustering, DBSCAN requires two parameter: ε and minPts. ε defines the radius of the searching area to count neighbors and minPts indicates the minimal number of points required for the center point to be classified as core point.

B. *Finding the optimum clustering result by multi-scale clustering process with DBSCAN*

As the city commercial activities usually show different aggregating characterizations at different scales, it is necessary to observe the activity cluster under different scales in order to find the optimum scale to describe the clusters of activities. The traditional DBSCAN is only able to detect clusters at given minPts and ε. Parameter ε decides the area of the unit searching circle, inside which clustering judgement would execute on points. Thus, the ε actually represents the minimum cluster area or the minimum clustering scale. In such case, as minPts can be determined, changing ε will allow researchers to discover the clustering result under different scales and furthermore, find the optimum ε for clustering the point set.

Taking a sample clustered point set as example, setting the minPts = 5 and let ε be a series of numbers from 0.2 to 2 with interval of 0.1. Except for the clustering results with different groups being colorized, the clustered group number curve against ε can be plotted. As can be seen from the result, until ε = 0.8 most points are kept un-clustered and the clustered group number keeps growing. When ε reaches 1, more points are getting assigned to clusters and the group number remains stable because only a large increase in ε could cover the gap between clusters. With ε increasing to 2, some close clusters in lower scales get combined into larger clusters and as a result, group number starts to drop down. Accordingly, considering the shape of the group number trend, it can be found that the clusters in this point set are most clearly clustered when the group number slides into stable segment from ε = 0.9 to ε = 1.4. Similarly, the curve become stable again from ε = 1.5 to ε = 1.8, where two separated but close clusters are merged. It can be inferred that as the ε keeps increasing, all the clusters will be merged into one cluster and the curve become stable again. These three stable segments indicate that all the clusters keep a distance from each other so that a small change of ε in between is not going to change the cluster dramatically. From the clustering results, the clusters of the point set are clearly grouped at stable segments.

In this way, as long as the global optimum minPts can be determined, it would be easy to figure out the optimum clustering result of a specific point set through a multi-scale clustering process.

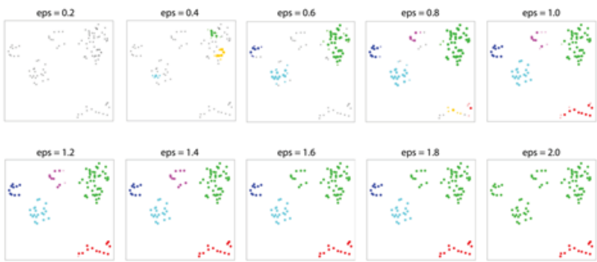

Fig. 1. Cluster results presented by different colors from eps =0.2 to eps = 2.0 under minPts = 5.

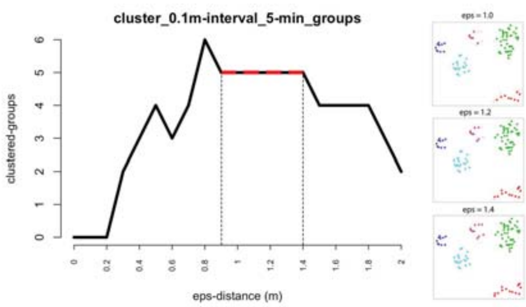

Fig. 2. The changing curve of total cluster number at different eps under minPts = 5.

C. *Determine global optimal minPts with modified DBSCAN algorithm*

The analysis on multi-scale clustering problem requires the minPts to be determined. First step is trying to locate the global optimum minPts in a qualitative way through diagrams based on the example above, then a programmable method is further introduced to derive global optimum minPts of any point set.

Following the above point set example and plotting two more cluster number curves at minPts =3 (yellow curve) and minPts = 8 (green curve) respectively, it can be seen that under different minPts, the maximum numbers of clusters are 13 at minPts = 3, 6 at minPts = 5 and 5 at minPts = 8. The maximum cluster number is decreasing as the minPts getting larger. This trend can be explained that the parameter minPts acts like a threshold of clustering in the algorithm. Another finding from the curves is that every two curves starts to overlap from the maximum cluster number of the curve with larger minPts. Based on this two findings, further step can be made to determine the global optimum minPts.

As mentioned above, when the cluster group number doesn't change dramatically with increased ε, the point clusters are clearly separated. In order to take different scales into consideration, all the stable segments need to be included in the clustering process. Thus, the maximum cluster number should stay as close as possible to the starting point(point E in figure 3) of the stable segments. Because if minPts is too large, some stable segments might be omitted and if minPts is too small, too many unnecessary clusters is going to be formed before reaching the stable segment.

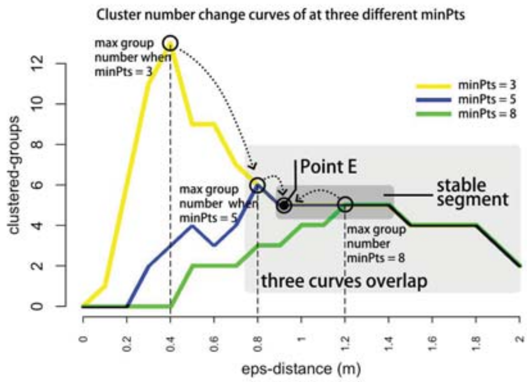

Fig. 3. Total cluster number curves underthree different minPts share the same stable segment and point E marks the starting point of the stable segment.

On figure 4, after plotting the changing curve of maximum cluster number under different minPts, it can be seen that the maximum cluster number keeps decreasing as minPts getting larger until cluster number becomes stable and the maximum cluster number will be equals to the same stable cluster number under different minPts. Thus, according to the approximate derivative of the line, the 'elbow' point A can be located easily, where the maximum cluster number stops dropping dramatically and begins to remain stable or decreasing slowly. The corresponding minPts 5 of the 'elbow' is classified as the global optimum minPts.

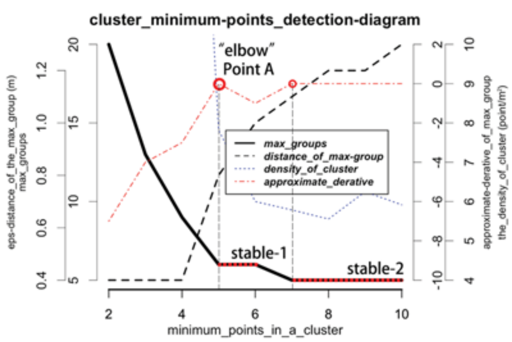

Fig. 4. Elbow point A of the maximum cluster number under minPts = 2 to 10 represents the global optimum minPts.

Based on the qualitative description of how to determine the global optimum minPts, a more programmable method is introduced below.

TABLE I. MODIFIED DBSCAN ALGORITHM

<table>
<tr><td>Input:</td><td colspan="2">A data set D containing 2-D points.</td></tr>
<tr><td rowspan="7">Variables:</td><td colspan="2">D: the data set.</td></tr>
<tr><td colspan="2">S: area of the point set space.</td></tr>
<tr><td colspan="2">n: number of points in D.</td></tr>
<tr><td colspan="2">ε: Searching radius for each point.</td></tr>
<tr><td colspan="2">minPts: the minimal points within range of eps for a point to be marked as a core point.</td></tr>
<tr><td colspan="2">C: a classified cluster.</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td>Step1:
Decide the range of ε</td><td colspan="2">Decide the range of ε as (0, Max_ ε] and the interval as C_interval.
The aim of this operation is to reduce unnecessary calculations and ensure that the majority of the points can be clustered at the Max_ε.
We recommend that:
Max_ε = 2*largest distance of 1-nearest-neighbor points.
C_interval = 1/20 median of the 1-nearest-neighbor distance.</td></tr>
<tr><td>Step2:
Decide the range of minPts</td><td colspan="2">Decide the range of minPts [2, Max_minPts]
We recommend that :
Max_minPts = ⌈n /S*Max_ ε^2* π ⌉.</td></tr>
<tr><td rowspan="5">Step3:
For each minPts in the range of loop DBSCAN with different to ε find the max cluster number</td><td colspan="2">For each minPts in the range, loop DBSCAN algorithm on different ε</td></tr>
<tr><td rowspan="3">Basic DBSCAN algorithm</td><td>Go through each points, count the points within its searching radius ε. Mark the points whose count > minPts as core point</td></tr>
<tr><td>For each core point, create a new cluster C as its classification. Check all the points within ε range and add them to cluster C if there is core points recursively do the check for all its connected core points.</td></tr>
<tr><td>Go through all non-core points and if it is reachable from any core point, put it into same cluster, otherwise mark it as noise point. Output result.</td></tr>
<tr><td colspan="2">Find the max cluster number under all possible ε for each minPts</td></tr>
<tr><td>Step4:
Decide global optimum minPts</td><td colspan="2">Plot the max cluster number curve for each minPts from Step 3. Calculate the approximate derivative of max cluster number line. Find the first minPts from which the derivative starts to be close to 0. We recommend to locate the first derivative that is less than 1/5 of the derivative range. This the corresponding minPts of this derivative is the global optimum minPts.</td></tr>
<tr><td>Step5:
Output the clustering results</td><td colspan="2">Execute the basic DBSCAN algorithm under all possible ε with the global optimum minPts. Plot the cluster number changing curve and the clustering result graphs for each ε.</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td>Output:</td><td colspan="2">Cluster number change curve under global optimum minPts
Clustering result of different ε under global optimum minPts</td></tr>
</table>

D. Find the local optimum ε and analyze the clustering results at global optimum minPts

While the global optimum minPts is determined, further analysis on clustering result under such minPts can be taken to evaluate the local optimum ε. Applying the modified DBSCAN to the above example point set, it is known that the global optimum minPts of the example point set is 5 and the cluster number change curve with clustering result under different ε is shown below in Figure 5. Based on the curve, scale-1, scale-2 and scale-3 are three stable region and thus three optimum clustering ε or scales could be further defined for this point set.

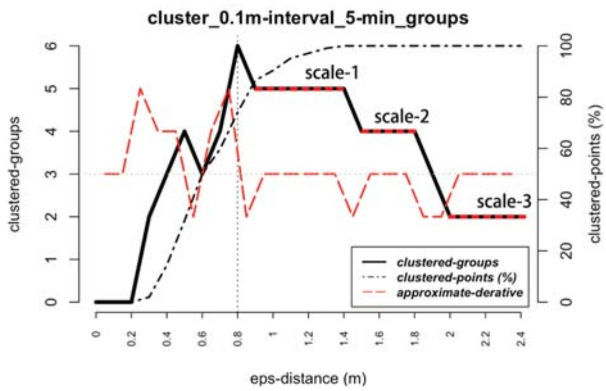

Fig. 5. The three stable segment at scale-1, scale-2, scale-3 on the cluster number curve present three local optimum ε under minPts=5.

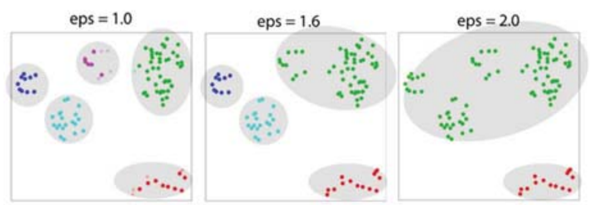

Fig. 6. Clustering results at three local optimum ε values detect clusters in different scales.

To sum up, the traditional DBSCAN can only detect clusters at the given minPts and ε. In this chapter, a new method is introduced to determine the global optimum minPts of DBSCAN automatically and, with global optimum minPts, further decide the local optimum ε. This modified DBSCAN can not only helping people to find the appropriate minPts and ε of a complex point set quickly, but also enabling the detection clusters under different scales.

IV. MILAN COMMERCIAL AGGLOMERATION ANALYSIS

In this part, the modified DBSCAN is applied to the commercial activities in Milan to analyse the spatial cluster of each commercial activity type. Combine the commercial data from two data sources and divide them into 25 commercial types. After all, four aspects of conclusion could be extracted from the clustering results.

A. Milan POI collection and classification

The POI data in this paper is mainly collected from Google Place API service. However the retail POIs from Google are not well classified, the registration data of fixed place retail from Milan Bureau of Industry and Commerce is used as substitute (Table II). Based on the ATECO 2007, which is the official category system used in Italy, all the POIs can be divided into 26 categories: accommodation, bar, restaurant, take away, food and drink shop, tabacchi (a kind of typical tobacco grocery store in Milan), supermarket, clothing, cultural service, cultural good store, car service, repair service, domestic good store, electronic store, pet and plant store, jewelry and watch store, cosmetic store, pharmacy, personal care and service, health and beauty service, financial service, real estate service, sport facility and fuel service.

TABLE II. POI DATA SOURCE

Source	Google POIs	Milan Bureau of Industry and Commerce
Data description	The POIs from Google place service.	The registration information of all retail
POI number	73440	25928
Update time	2017	2018
Category system	Google place types[a]	ATECO 2007[b]
Categories number	88	179

[a] The detail types can be found on https://developers.google.com/places/webservice/supported_types#table1

[b] ATECO 2007 is based on NACE. The complete category can be found on https://www.istat.it/en/archivio/17959

B. *Milan commercial agglomeration analysis*

For each commercial activity type, the modified DBSCAN algorithm is able to return a cluster number curve through a range of ε at the global optimum minPts (Figure 7, 8) and a series of clustering result pictures with different colors representing different groups (Figure 9). Here in table II, the global optimum minPts, local optimum ε and the estimated density for each commercial activity type are listed. Based on clustering results at the optimum minPts and ε, the conclusions from 4 aspects can be further drawn.

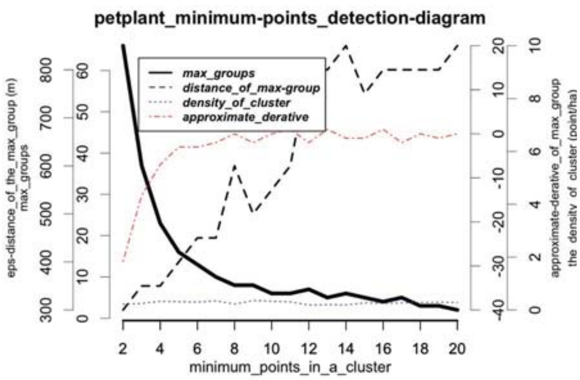

Fig. 7. Pet and plant store activity max cluster number curve at different minPts.

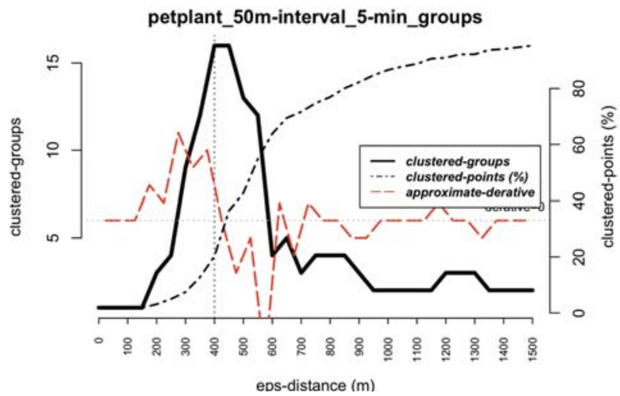

Fig. 8. Pet and plant store activity cluster number curve under the global optimum minPts.

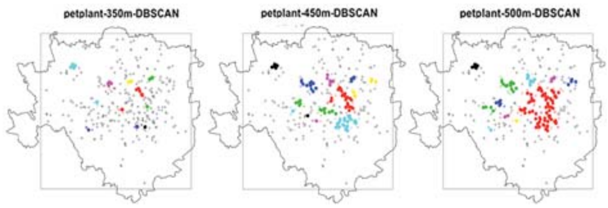

Fig. 9. Pet and plant store activity cluster at three different ε values under the global optimum minPts.

The local optimum ε can be decided by finding the 'stable' section of the clustered group number curve and the approximate density of the agglomeration can be calculated with minPts and optimum ε. The results of all commercial types are listed below in the Table II.

TABLE III. CLUSTERING RESULTS

	Optimum minPts	Optimum eps 1 (m)	Optimum eps 2 (m)	Cluster unit density (p/ha)
accommodation	9	150	220	1.27
bar	10	120	200	2.21
beauty	7	130	200	1.31
car	7	240	350	0.39
clothes	18	130	150	2.92
cosmetic	7	250	400	0.36
culturalgoods	7	130	240	1.32
culture	5	130	200	0.94
Domestic goods	7	150	240	0.99
electronics	6	240	380	0.33
financial	7	190	250	0.62
food	11	180	240	1.08
fuel	4	600	900	0.04
health	10	170	240	1.01
jewelry	4	150	340	0.57
leisure	5	450	600	0.1
Pet plant	5	450	500	0.1
pharmacy	5	400	500	0.1
Real estate	7	240	320	0.39
repair	4	270	500	0.17
restaurant	8	130	200	1.51
sport	4	260	400	0.19
supermarket	4	650	1000	0.03
tabacchi	5	350	700	0.1
takeaway	8	300	400	0.28

1) *Agglomeration scale analysis*

From the table above, it can be seen that different commercial types usually have different first local optimum ε, which means the different types of activities tend to cluster at different scales. Comparing all the first local optimum ε, it can be found that the supermarket and fuel activity have the largest unit cluster radius(more than 500m), while the bars, clothing stores, personal care ,beauty shops, et al. could be able to cluster at around 100m. The figure 10 below classifies all local optimum ε into 5 range categories and illustrate the unit size of the cluster with a circle in the map of Milan. Food and drink shops, health facilities, jewelry stores and financial activities form stable aggregations at ε = 200m. Real estate agencies, cosmetic stores, sport facilities, electronic stores and et al. cluster ε = 300m. Pet and plant shops, leisure places and pharmacies have second largest unit cluster radius 400m.

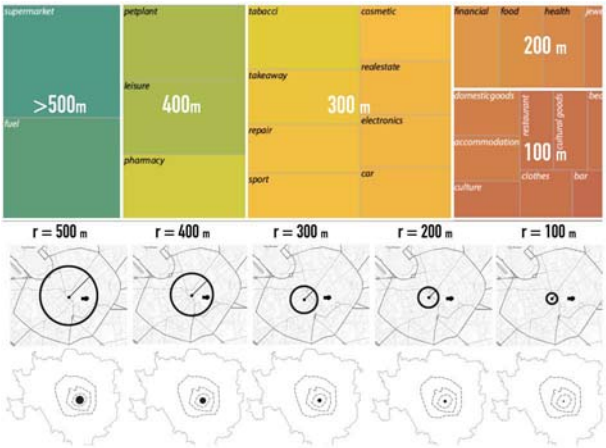

Fig. 10. Unit cluster size of different types commercial activity and the comparision between unit cluster size and Milan inner ring road.

2) Agglomeration density analysis

Another important feature to describe the agglomerations is the activity density inside the cluster. According to the global optimum minPts and local optimum ε, a density threshold can be calculated using the following function:

$$\rho = \text{global optimum minPts} / (\text{local optimum } \varepsilon)^2 / \pi$$

Plotting the density threshold in a descending order, it can be found that the unit density of clothing store cluster is the highest(around 3 stores per hectare), followed by the density of bars with 2.2 / hectare and restaurants with 1.5 / ha. The average threshold density of all commercial activities is 0.7 / ha. Leisure places, pet and plant shops, pharmacies and so on have the lowest density.

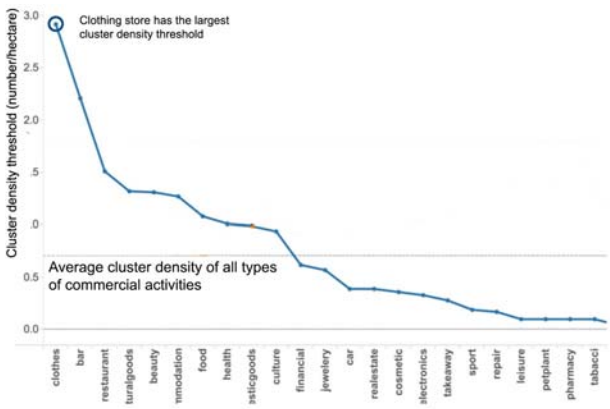

Fig. 11. Cluster density of each type of commercial activity in Milan.

3) Agglomeration shape analysis

In addition to aggregation scale and aggregation density, through visualization of the clusters, it can be seen that different retail and service types have their own shape of patterns. A typical example is clothing stores, in Milan they usually extend along the streets, such as Corso Buenos Aires, Via Torino, and Corso Vercelli in the figure 12 below. On the contrary, some other activities are clustered in roughly square shape. For example, the accommodations, like hotel and hostel, aggregate around Milan Central Railway Station, Cinque Vie and so on in square shape.

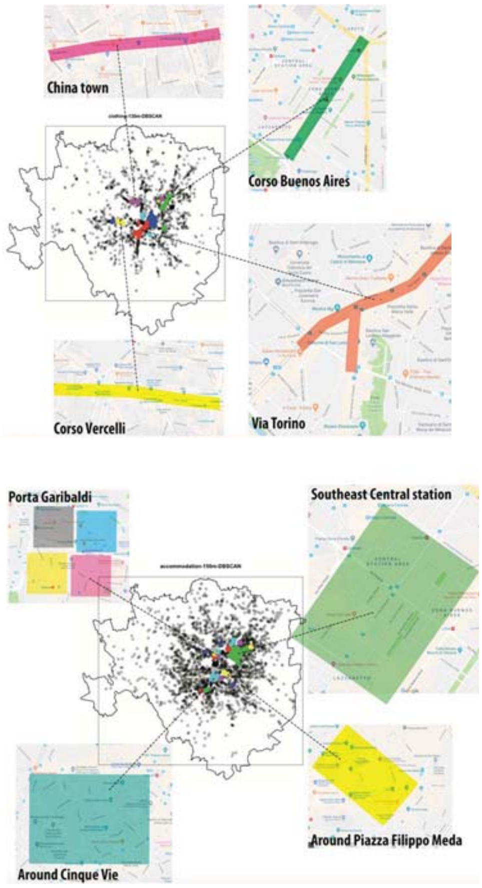

Fig. 12. Linear-shaped clothing store clusters and planar-shaped accommodation clusters in Milan.

4) Agglomeration spatial distribution pattern analysis

Different types of commercial activities are not only differ greatly in the density, scale and shape of clusters, but also show different patterns in the overall spatial structure. This paper identifies the spatial location of clusters and summarizes five commercial agglomeration spatial structure patterns. (1) Central aggregation pattern: Majority of the activities are concentrated in the central area of Milan, such as culture related activities, jewelry stores and so on. (2) Ring around center pattern: Activities like electronic stores, car-related services are distributed around the inner ring or outer ring road in Milan. (3) High-density aggregation pattern: the activities in this pattern are highly concentrated in a certain area of Milan, typically represented by financial services activities. (4) Disperse distribution pattern: the clusters of his commercial activity type spread all over the city and regularly distributed in Milan, such as sports facilities, food and drink shops, personal caring and beauty services, pet and plant shops. (5) Hierarchical distribution pattern: commercial activities in this pattern show an obvious hierarchical relationship in cluster size and center, sub-center clusters can be further identified according to the size of the clusters. Bars, clothing stores, restaurants belong to this type of pattern.

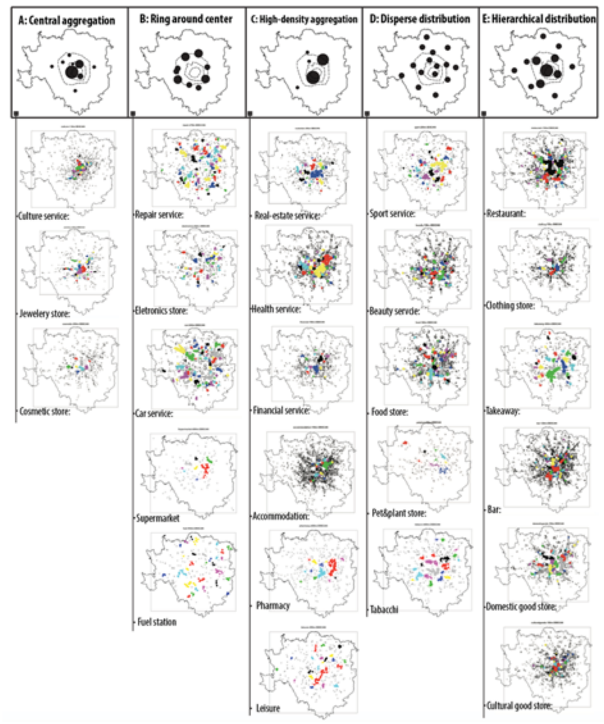

Fig. 13. Five commerical activity cluster spaital patterns in Milan. A: central aggregation pattern, B: Ring around center pattern, C: High-density aggregation pattern, D: Disperse distribution pattern, E: Hierarchical distribution pattern.

V. Conclusion

Different types of commercial activities in a city usually aggregate at different scales. Thus, in order to grasp the aggregating features for each commercial activity type, we need to examine the POIs in a range of scales to find out one or several best suitable observation scales to identify the clusters. To solve this problem, a modified DBSCAN algorithm that can be used to find the global optimum minPts automatically and locate the local optimum ε based on the shape of cluster number change curve under global optimum minPts is introduced. The benefit of this modified DBSCAN algorithm, compared to its original algorithm, is that it frees the researchers from practicing whole lot of parameters to get a high performance result and it is able to search the optimal global minPt and the optimal ε under such minPt automatically. The scope of this algorithm applies the best to the problems that involves in multiple heterogeneous point set, with possible cluster under variety of density. This modified algorithm aims to identify the most representable clustering result of a complex point set with clusters at different scales. Furthermore, the new algorithm is applied on the 25 types of commercial activities in Milan and based on the results, conclusions were drawn from four aspects: agglomeration scale, agglomeration density, agglomeration shape and agglomeration structures. It can be confirmed that the modified DBSCAN is capable of detecting the clusters at different combination of ε and minPts and further comparing the results to find the optimum one. Furthermore, the modified DBSCAN algorithm could be refined and developed in many directions. New topics including but not limited to how to expand the method of finding the global optimum minPts in a more quantitative way and how to select the local optimum ε in a more automatic favor. We hope to continue the work in the future.

References

[1] M. E. Porter, “Clusters and the new economics of competition,” Harv. Bus. Rev., vol. 76, no. 6, pp. 77-+, 1998.

[2] Glaeser, “Consumer City,” vol. 1, no. 1901, pp. 27–50, 2000.

[3] M. F. Goodchild, “Citizens as sensors: The world of volunteered geography,” GeoJournal. 2007.

[4] S. Jiang, A. Alves, F. Rodrigues, J. Ferreira, and F. C. Pereira, “Mining point-of-interest data from social networks for urban land use classification and disaggregation,” Comput. Environ. Urban Syst., vol. 53, pp. 36–46, 2015.

[5] W. Chen, L. Liu, and Y. Liang, “Retail center recognition and spatial aggregating feature analysis of retail formats in Guangzhou based on POI data,” Geogr. Res., vol. 35, no. 4, pp. 703–716, 2016.

[6] W. Yu and T. Ai, “The visualization and analysis of POI features under network space supported by kernel density estimation,” Cehui Xuebao/Acta Geod. Cartogr. Sin., 2015.

[7] J. L. Dungan et al., “A balanced view of scale in spatial statistical analysis,” Ecography (Cop.)., 2002.

[8] M. J. Crawley and J. E. Harral, “Scale dependence in plant biodiversity,” Science (80-.)., 2001.

[9] G. Sheikholeslami, S. Chatterjee, and A. Zhang, “WaveCluster : A Multi-Resolution Clustering Approach for Very Large Spatial Databases,” in The VLDB Journal The International Journal on Very Large Data Bases, 1998.

[10] L. Ertöz, M. Steinbach, and V. Kumar, “Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data,” 2013.

[11] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in Kdd, 1996.

[12] T. Czerniawski, B. Sankaran, M. Nahangi, C. Haas, and F. Leite, “6D DBSCAN-based segmentation of building point clouds for planar object classification,” Autom. Constr., 2018.

[13] D. R. Edla and P. K. Jana, “A Prototype-Based Modified DBSCAN for Gene Clustering,” Procedia Technol., 2012.

[14] X. Shou-kun, W. Chao, Z. Li-hua, and G. Xin-hua, “DBSCAN Clustering Algorithm for the Detection of Nearby Open Clusters Based on Gaia-DR2two,” Chinese Astron. Astrophys., 2019.

[15] K. Khan, S. U. Rehman, K. Aziz, S. Fong, S. Sarasvady, and A. Vishwa, “DBSCAN: Past, present and future,” 5th Int. Conf. Appl. Digit. Inf. Web Technol. ICADIWT 2014, pp. 232–238, 2014.

[16] M. Pavlis, L. Dolega, and A. Singleton, “A Modified DBSCAN Clustering Method to Estimate Retail Center Extent,” *Geogr. Anal.*, 2018.

[17] P. Liu, D. Zhou, and N. Wu, “VDBSCAN: varied density based spatial clustering of applications with noise,” in 2007 International conference on service systems and service management, 2007, pp. 1–4.

[18] M. Naik Gaonkar and K. Sawant, “AutoEpsDBSCAN : DBSCAN with Eps Automatic for Large Dataset,” Issn, no. 22, pp. 2319–2526, 2013.

[19] X. Liu, Q. Huang, and S. Gao, “Exploring the uncertainty of activity zone detection using digital footprints with multi-scaled DBSCAN,” *Int. J. Geogr. Inf. Sci.*, 2019.