Identifying Micro Trade Areas Using Sales Information

Yong Li, Han Ren

Department of Automation, Kunming University of Science and Technology, Kunming, Yunnan, 650550, China

Corresponding author is Yong Li

Abstract
The delineating methods of traditional trade areas are based on location information, and can not reflect changes in the point of trade areas, affect the launch of products and other marketing decisions in the situation that massive mobile terminals are used. Based on the analysis of investigation and survey technology on traditional trade area, this paper proposes a method to indentify or delineate trade areas based on big data analytics by focusing on the correlation and real-time data of trade areas, which we name it micro trade area partition (MTA partition). By using about 700 retail stores' data of Guiyang Tobacco Company from Jan to Mar in 2014, the valid of the proposed method has been verified.
Keywords: MARKETING DECISION, MICRO TRADE AREAS, BIG DATA ANALYSIS, IDENTIFICATION.

1. Introduction

Trade area analysis is an important topic of marketing campaign to successfully maintain companies' continuous growth and survival, and it is very important to guide the rationalization layout and construct retail and service networks at a specific geographic location as well as making sales estimates and determining a future marketing strategy [3-5]. Normally, trade area is the geographic area from which the store draws most of its customers and within market penetration is highest [12]. But with the massive use of mobile phone, the behavior of customers are changed rapidly, especially with the development of electronic commerce, customers are more and more using mobile phone or other digital terminals to buy their needs instead of going to retail stores, and pay little care about of whatever the stores geographic positions are. In this case, the use of APP technology can help firms or personals implementing their business on mobile terminals, accomplishing network marketing, and constructing the companies or personals' new trade area. Compared to the traditional trade area, the use of APP such as Wechat enables businesses to form a new kind of trade area, which is call micro trade area [20].

The focus of traditional trade area is the geographical distribution of customers and retail stores. In traditional classification, the trade area normally indentified as governmental district, school district, commercial mall district, tourists attractions and so on. In many cases, the categories of these districts could not reflect differences among stores in the same area, and its direct result is the same delivery strategy to be used in different retail stores, and could not obtain better practice in sales and marketing strategic planning for a company. There is little discussion on micro trade areas. In traditional theory, several researchers describe trade areas as an what marked by various qualitative indicators relative to the surrounding urban environment as well as intrinsic marketing factors of the retail store, i.e. attractiveness, prices, size of the outlet, diversity of the merchandise [11].

But with the use of APP, the limitation of location has broken down, especially accompany with the development of logistics, the customers may be far away from stores and not live in a specific geographic location. In the situation, the size of the stores, environmental factors are not the key points that restrict the sales of stores. Thus, there are obviously differences between traditional trade area and the micro trade areas, such as definition and usage mode. The traditional partition method of trade area based on the geographic information theory could not be strictly applied to retail stores.

The notion of "Big Data" is now being recognized broadly in the field of data-driven decision-making [1]. There is growing enthusiasm for the value of data, and products obtained through analyzing it. According to Gartner's definition, Big Data is high-volume, high-variety and high-velocity information assets. The data of retail stores are very large, various structures, and time-dependence. It is obviously that the retail data conforms to the definition of Big Data. The corresponding Big Data methods also could be used to analysis micro trade areas.

This paper proposes a method called Micro Trade Area (MTA) partition to identifying trade area based on Big Data analytics by focusing on the correlation and real-time data of retail stores to improve the application value of trade area analysis. It provides a basis for understanding, quantifying, and visualizing customer flow and movement about a retail store in a new business model. And a practical case of cigarette retail stores in Guiyang city of China are taken as an example to verify the proposed method at last.

2. Literature Review

There are three types of methods used for delineating trade areas, i.e. theoretical, empirical, and statistical methods. The most common theoretical methods are the central place theory of Christaller and retail gravity models [13]. According to the closest center hypothesis, the proximal area method assumes that customers will choose the nearest stores to them, which are polygon areas closer to a store than to any other stores [14]. The retail gravitation models was proposed by Reilly, and extended as the Huff model and the MCI model to delineate geographical division of the store market area, in which the size of store and the distance between the home and the store were concerned [15, 17].

The driving time method and the analog method are two typical empirical methods by many practitioners. The driving time method has indeed numerically proved that the driving time required to reach a set of outlets is highly influential in determining consumer shopping center choice [18]. The most popular empirical method for determining trade area borders is the analog method. It delineates the size, shape and character of stores' trade area by customer addresses [19].

The statistical methods mainly contain the regression method and the clustering methods. The regression method seeks to measure a parameter of performance for estimating the market share of zones surrounding a new outlet location for delineating trade areas by correlating it with various socio-economic, environmental and marketing variables. The clustering methods has been used to specify trade areas limits of stores with the implementation of functions which consist in minimizing the curve radius in each point of the trade area border function to make it more regular [16]. All these methods are often based on the geographical information even if some available data coming from experience can be used.

For complexity, scale, timeliness with Big Data, there are many challenges during the process from transforming Big Data into a structured format to later analysis with these data [7-9]. Here we only discuss the analysis methods with Big Data on identifying micro trade areas. For any enterprise, knowing its retail stores whether they could compensate for the requirements of changing expectations of the city dwellers or other people from trade areas becomes a key point that effect the marketing policy [10]. In traditional retail trade area analysis, consumer flow and direction might change due to newly opened or closed stores and the changing composition of the local population, regular retail trade area analysis is necessary to track a store's current location and customer flow. But with the use of mobile phone, the current location of retail stores is no more the only key of trade area. The customer flow can be still reflected from sales data and amount, thus, using these data for analysis the trade area of stores is reasonable. In summary, Big Data makes the academic community to face the new opportunities and challenges [2]. And using Big Data analysis approach to solve business problems has become a trend, such as consumption forecasts [6].

3. Methodology

3.1. Limitations of the traditional division of trade area

In the analysis of traditional partition trade areas, two kinds of data were typically used which collected from the business segments surveys as follows:

1) Commercial environmental data. The data were the information of the dominative areas of neighborhood communities, residential districts, villages, major government agencies, organizations, schools,

competition retail shops, farmers' markets, restaurants and traffic conditions, which collected from the covered areas with center at a certain retail store and of radiuses 1 km, 3 km and 5 km respectively.

2) Passengers data. The data are the number of households, population, employed persons, temporary habitant population and their roughly income level in the commercial environment mention above. It also contained the number of dine together and the past shopping destination of the persons in commercial environment, especially the persons of major government agencies, organizations and schools.

By summarizing and classifying the two kinds of data, we could find:

Firstly, collecting data were in order to analyze the causal relationship of these data from trade areas. For example, there are causal relationships among consumption levels, the number of passengers, commodity sales and grades. But it was difficult to be quantified in practical application.

Secondly, it would be difficult to fully collect the two kinds of data. Although the trade areas have been identified as governmental district and school district, the different between two neighborhood cigarette retail stores in the same district could not be differentiated. If trade areas were further identification, the increased workload and difficulty of implementing with geometric growth level involved. In our study, the traditional method for identifying trade areas also cannot meet the practical needs in cigarette industry.

In order to verify the points mentioned above, we randomly selected cigarette retail stores in a district of Guiyang city for analysis. The data are collected from V3 software system used in Guiyang Tobacco Company, containing 636 samples, which 10 samples in the dataset are listed as in Table 1.

Table 1. Sales Data of Retail Stores from Jan. to Mar. in 2014

No.	X_1	X_2	X_3	X_4	X_5	X_6
01	1245	191970	435	62319	290	31823
02	180	10226	35	2658	90	4790
05	945	112438	435	54567	470	57811
09	915	126167	85	13643	500	69736
10	1160	120126	300	30233	340	39265
10	895	133650	250	34202	600	90365
10	1345	194302	580	76382	600	75189
08	850	135320	275	37714	350	47453
06	260	21347	250	19741	180	14697
10	605	101593	115	14865	235	24241

Among the data items, No. labels uniquely a retail store related with geographic location. Sales contain sales volume and sales amount. We presupposed that the existed zones are reasonable, and it reflected the difference of sales and amount in the same area. Based on this assumption, we designed a linear regression model to fit the collected data in Table 1.

The regression equation is as follows:

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \varepsilon, n = 6 \tag{1}$$

Y denotes existed districts. X_1, X_2, X_3, X_4, X_5, X_6 represent sales volume and amount of retail stores from January to March in 2014. ε describes the combined effects on the dependent variable Y of effects of minor factors, random change, measurement error and model set errors of mathematical form and other factors, commonly referred to as stochastic disturbance. Constants β_i are undetermined parameters. We select the trade areas as the dependent variable, the sales and amount as the independent variables, and directly use sales data of 3 months to evaluate goodness of fit of the formula (1). We use SPSS (Statistical Product and Service Solutions, Version 19.0) for evaluation [11]. The results are listed in Table 2, Table 3, and Table 4.

Table 2. Model Summary of Linear Regression Test

Model	R	R^2	Adj R^2	Std. error of estimate
1	.076[a]	.006	-.004	2.852

[a] Predictors: Constant, sales volume and amount.

Table 3. The Anova of Linear Regression Test with Model 1

	Anova[b]				
	Sum of squares	Df	Mean square	F	Sig.
Regression	29.593	6	4.932	.618	.716[a]
Residuals	5036.7	631	7.982		
Total	5066.3	637			

[a] Predictors: Constant, sales volume and amount.
[b] Dependent variable: No. of existed trade area.

From Table 2 to Table 4, we note that:

1) R^2=0.006, it points out that 0.8% of the variation of existed trade areas can be explained by six variables, or the six variables can explain 0.8% of the variation of existed trade area.

2) The values of Sig are high. It indicates that assumption is not significant, i.e., the differences among existed trade areas are not more obviously, that means the difference can’t present different cigarette sales.

Table 4. The Coefficient of Linear Regression Test with Model 1

	Non std. Coef.		Std. coef.	t	Sig.
	β	Std. error	Trial version		
Constant	10.655	.577		18.465	.000
X_1	-.005	.004	-.315	-1.293	.197
X_2	.000	.000	.441	1.708	.088
X_3	.021	.008	.490	2.442	.015
X_4	.000	.000	-.558	-2.505	.013
X_5	-.010	.008	-.294	-1.345	.179
X_6	.000	.000	.328	1.413	.158

* Dependent variable: No. of existed trade area.

This conclusion confirms the opinions that the existed trade areas based on geographic partition could not satisfy the practical application. The idea of identifying the trade area or MTA partition of cigarette retail stores with Big Data method is as following:

Firstly, we use the real stability data and dynamic real-time data that can be collected for analysis. Secondly, we focus on considering data correlation, regardless of causality. Thirdly, we use the most direct result data for prediction and data analysis.

Based on the three steps mention above, we use Big Data analysis for MTA partition with the existing sales volume, sales amount of cigarettes, and cigarette structure data of Guiyang Tobacco Company. These data is accumulated for long time and specific to each cigarette retail stores.

3.2. Data Preparation of MTA partition

We take the data of Zone 1 in Table 1 as an example for the MTA partition, which contains 42 retailers' data, and 10 samples in the dataset are listed as in the Table 5.

Table 5. Sales Data of 10 Retail Stores from January to March in 2014 in Zone 01

No.	X_1	X_2	X_3	X_4	X_5	X_6
31	1245	191970	435	62319	290	31823
69	1970	323875	465	92450	485	87044
77	1155	123525	320	30142	230	22809
87	380	36198	185	18321	415	39676
89	1055	145536	675	80331	720	94153
107	295	30436	325	25382	140	9142
135	1170	203118	295	50639	335	46372
151	80	7803	55	5480	95	10994
159	475	70222	335	50372	325	52025
160	495	67139	340	45688	455	52280

* No: Number of retail stores in Zone 1.

For further discussion, we need to predispose the data as in Table 6 following next two steps:

1) Adding data: add cigarette structure data for analysis. Cigarette sales data reflect quantity and cigarette sales structure data reflect quality of the sales level. The amount of cigarette sales is a combination of the sales volume and sales structure:

Structure = Amount/Volume (2)

Based on formula (2), we could obtain the prepared three-month data of retail stores.

2) Standardizing data: the measurement units of three data types, i.e., cigarette volume, amount and structure are different, thus numerical methods must be used to standardize all variables. We use standard deviation to standardize the variables (z-score standardization) marked as Zscore.

Table 6. Standardized Variables of Zone 1

No.	Z_1	Z_3	Z_3
1	.73	.58	.19
2	1.50	1.73	1.28
3	.39	-.03	-.66
4	-.10	-.33	-.75
5	1.61	1.04	.05
6	-.27	-.51	-1.09
7	.49	.61	.81
8	-.95	-.76	-.56
9	.11	.18	.61
10	.30	.13	.01

*Z_1: Zscore (Average of volume), Z_2: Zscore (Average of amount), Z_3: Zscore (Average of structure).

3.3. Using Clustering Algorithm to Identify Trade Area in MTA partition

After the data is ready, we use clustering algorithm to identify trade areas.

1) Choosing identified trade area: taking Zone 1 as an example (contains 42 samples), delineating trade areas. We select the hierarchical clustering to form a tree intuitively statement classification of the certain zone as shown in Figure 1. From Figure 1, the zone is identified as two regions, the No. 383 retail store as a separate category, others retail stores as another category in the first layer of clustering. In the second layer of clustering, the zone is identified as three, up to 8 categories. Here we take three categories as illustrative example.

2) Selecting cluster variables and methods: taking cigarette volume, amount, structure as variables, using a hierarchical clustering method and k-means clustering method for analysis. Based on hierarchical clustering, we use k-Means clustering method to identify the trade area, preset classification as three

categories, and obtained 8 retail stores in category 1, 45 retail stores in category 2 and 1 retail store in category 3 as in Table 7.

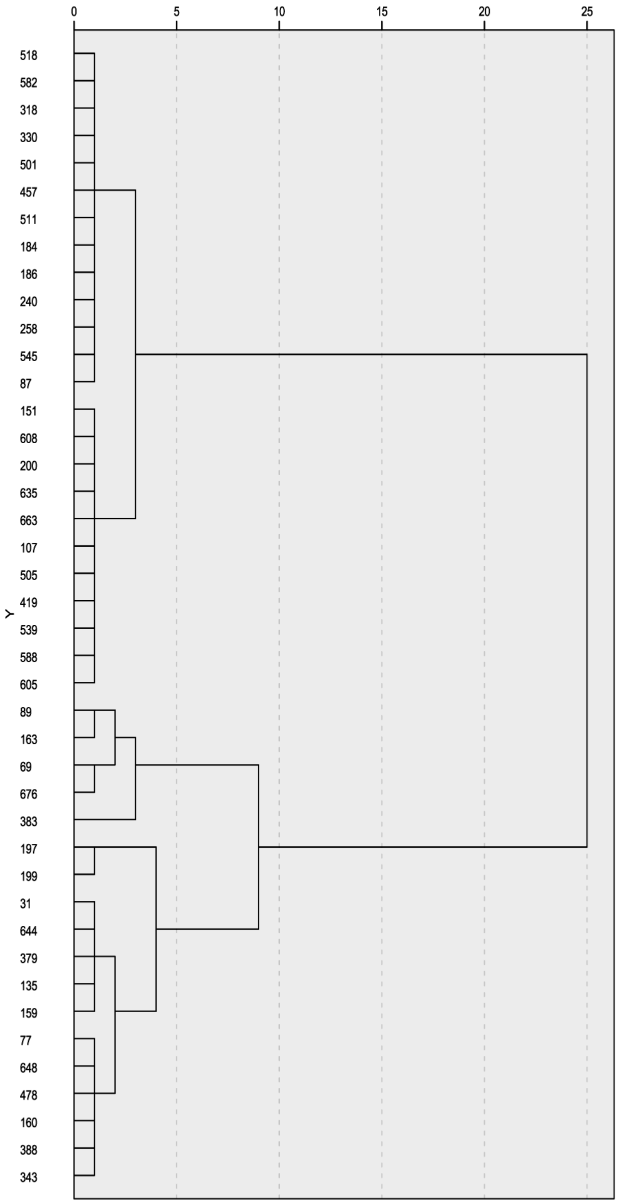

Figure1. The hierarchical clustering tree of Zone 1

Table 7. The Categories of Identified Trade Areas

No. Number of stores		
Cluster	1	1
	2	14
	3	27
Effective Cases	42	
Dismissed Cases	0	

In summary, the results we obtained are consistent.

Compared with traditional identifying method of trade area, the MTA partition method with clustering sales data has the following advantages:

1) Directly using the existing sales data, reducing the cost of data collection.

2) Identifying trade areas more detailed. For example, in Zone 1, hierarchical clustering method can identify 54 retail stores up to 9 categories, even meticulous to identify a retail store as a category of trade area (No. 383). Of course, in the application of identifying trade area classification, it needs to consider the value that identifying a retail store as a classification of trade area in a practical sense.

3) MTA partition does not simply stick to geographical location, but its' results fully associated with cigarette sales, so it can be applied easily and can avoid the problem in traditional delineating method.

In summary, the 42 retail stores are belong to the same category named Zone 1, and the marketing strategy of cigarette supplying is also same. It can't achieve the marketing goal that spending minimum total cost for obtaining maximum total profit from either the manufacturers or retailers' side. The MTA partition can support the manufacturers or retailers adopting more marketing strategies to serve the customers in a certain trade area.

3.4. The Validation of the MTA Partition

To ensure the consistency of the results, we use curve fitting method to verify the goodness of fit of the MTA partition. Using SPSS, we obtain the linear regression result as in Table 8, 9 and 10.

Table 8. The Model Summary of Linear Regression Test on the MTA Partition

Model	R	R^2	Adj R^2	Std. error of estimate
1	.883[a]	.780	.762	.263

[a] Predictors: Constant, Z_1, Z_2, Z_3.

Table 9. The Anova of Linear Regression Test on the MTA Partition with Model 1

	Anova[b]				
Model [c]	Sum of squares	Df	Mean square	F	Sig.
Regression	9.284	3	3.095	44.876	.0[a]
Residuals	2.621	38	.069		
Total	11.905	41			

[b] Dependent variable: Category No. of cases.
[c] Predictors: Constant, Z_1, Z_2, Z_3.

From Table 8 to Table 9, we note that:

1) R^2=0.78, it points out that 78% of the variation of 3 categories in Zone 1 can be explained by three sub-indexes, or the three sub-indicators can explain 78% of the variation of subdivided trade area.

Table 10. The Coefficient of Linear Regression Test on the MTA Partition with Model 1

Non std. Coef.			Std. coef.	t	Sig.
	β	Std. error	Trial version		
Constant	2.526	.044		57.342	.000
Z_1	-.438	.234	-.763	-1.873	.067
Z_2	.121	.344	.162	.351	.727
Z_3	-.358	.108	-.466	-3.323	.002

*Dependent variable: Category No. of cases.

2) The values of Sig are low. It indicates that assumptions are significant, i.e., by identifying the trade area, the differences between the different trade areas are more obvious, that means the difference presents different cigarette sales.

From the Table 10, the regression equation can be obtained as following:

$$Y = 2.526 - 0.438X_1 + 0.121X_2 - 0.358X_3 \quad (3)$$

In the formula (3), Y represents the classification, X_1 denotes cigarette sales, X_2 is on behalf of cigarette sales amount, and X_3 represents the cigarette sales structure. Among them, the sales and amount of cigarette sales impact on the classification are -0.438 and -0.358 (negative correlation), greater than the impact of cigarette structure (0.121), then, in practical sense, the impact of cigarette structure can be ignored.

The characteristics of the three micro trade areas in Zone 1 can be determined:

Category 1: the amount of sales is relatively high, but a few sales volume. These retail stores belong to high-end trade area;

Category 2: the sales amount and sales are at comparable level. These retail stores belong to the mid-end trade area, and are the most in all stores;

Category 3: the amount of sales is relatively small, and also is sales. These stores belong to low-end trade area.

We use the MTA partition method to delineate one certain area, and we can delineate the others in the same way. The result is more detailed than the traditional classification, thus MTA partition can give more support for decision maker.

4. Conclusion

Based on the traditional method for delineating trade area, we took the practical sales data of Guiyang Tobacco Company as an example, proposed a method called the MTA partition with clustering algorithm to identify trade area using Big Data analysis. By validating the proposed method, we concluded that the categories of trade area are on a high degree of correlation with cigarette sales, cigarette sales amount, cigarette sales structure three data types. Compared with traditional delineating trade area methods, the method has more significant advantages in the judgment of retail stores categories, prediction on cigarette sales and structure, supply delivery and other areas in practical sense. It should be pointed out that the data types that have been used in the MTA partition method are relatively few. With the development in data collection methods, using more types of practical data, and increasing the independent variable in delineating trade area, the proposed method should have further actual application results.

References

1. D. Agrawal, P. Bernstein and E. Bertino, et al. Challenges and Opportunities with Big Data – A community white paper developed by leading researchers across the United States, http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf, 2012.
2. X. F. Meng, X. Ci. Big data management: concepts, techniques and challenges. Journal of Computer Research and Development, 2013, 50(1), pp.146-169.
3. James J. X. Wang, J. Xu, An unplanned trade area in a fast-growing city: a case study of Shenzhen, China. Journal of Retailing and Consumer Services, 2002, 9(6), pp.317-326.
4. M. Kakoi, S. Saito, Extracting rules of loyal visitors' characteristics at central trade areas based on on-site sampling. Studies in Regional Science, 2006, 36(4), pp.993-1005.
5. M. Angélica Salazar-Aguilar, Roger Z. Ríos-Mercado, José L. González-Velarde. A bi-objective programming model for designing compact and balanced territories in trade area-ing, Transportation Research Part C: Emerging Technologies, 2011, 19(5), pp.885-895.
6. N. Marz, J. Warren. Big Data: Principles and Best Practices of Scalable Real-time Data Systems, Manning Publications, 2012, 328 p.
7. J. Manyika, M. Chui. Big Data: The next frontier for innovation, competition and productivity, McKinsey Quarterly, 2011, 5, pp.27-30.
8. N. Archak, A. Ghose. Deriving the pricing power of product features by mining consumer reviews, Management Science, 2011, 57(8), pp.1485-1509.
9. J. Howe. Crowd sourcing: How the Power of the Crowd is Driving the Future of Business, Crown Publishing Group, 2006, 320 p.

10. S. E. Kilic. Preservation plan applications for the historical city centre, Kemeralti (Izmir, Turkey), European Planning Studies, 2008, 16(2), pp.253-276.
11. J. Pallant. SPSS Survival Manual: a step by step guide to data analysis using IBM SPSS, Open University Press, 2013, 368 p.
12. A. Ghosh, S. L. McLafferty. Location Strategies for Retail and Service Firms, Rowman & Littlefield, 1987, 224 p.
13. Je´ro^me Baray, Ge´rard Cliquet. Delineating store trade areas through morphological analysis, European Journal of Operational Research, 2007, 182, pp.886–898.
14. A. Getis. The determination of the location of retail activities with the use of a map transformation, Economic Geography, 1963, 3(1), pp.14–22.
15. W. J. Reilly. The Law of Retail Gravitation, W.J. Reilly, 1931, 75 p.
16. D. L. Huff, R. R. Batsell. Delimiting the areal extent of a market area, Journal of Marketing Research, 1977, 14(4), pp.581–585.
17. M. Nakanishi, L. G. Cooper. Parameter estimation for a multiplicative competitive interaction model – least square approach, Journal of Marketing Research, 1974, 11 (3), pp.303–311.
18. J. A. Brunner, J. L. Mason. The influence of driving time upon shopping center preference, Journal of Marketing, 1968, 32 (2), pp.57–61.
19. W. Applebaum. Methods for Determining Store Trade Areas, Market Penetration, and Potential Sales, Journal of Marketing Research, 1966, 3(2), pp.127–141.
20. Wiki. https://en.wikipedia.org/wiki/WeChat, 2015.

The Influence of Insider Law on preventing insider information leakage

Guangshuai Li, Ying Bi

School of Economics and Management, Beijing Jiaotong University, Beijing, 100044, China

Corresponding author is Guangshuai Li

Abstract
This paper analyzes the inhibition mechanism of the insider law based on the legislation and enforcement using the data of 50 countries (regions).Our findings indicate that legislation is a more important factor affecting insider information leakage. The perfection of legislation can give authority more power to improve the protection for investors. On the other hand, it can help establish a strict information disclosure system for the listing corporation and the intermediary organizations, in which way the leakage of insider information is prevented. In addition, some suggestions about regulatory policies and legal provisions in stock market are made in this paper.
Keywords: LEGAL SYSTEM, LEGISLATION, EXECUTION, INSIDER INFORMATION LEAKAGE.