Subdividing Trade Area of Cigarette Retail Stores Based on Big Data Analytics

Li Yong, Li Qianye

Institute of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650550，China
E-mail: Dr.Leon@foxmail.com

Abstract: The traditional partition method of trade area based on location information can not reflect changes of trade area, affects the launch of products and other marketing decisions with use of massive mobile terminals. Based on the analysis of the traditional partition method on trade area, this paper proposes a method to subdivide trade area based on big data analytics by focusing on the correlation and real-time data of trade area, which we name it micro trade area partition (MTA partition). By using about 700 retail stores' data of Guiyang Tobacco Company in a certain area from Jan to Mar in 2014, the validity of the proposed method has been verified. The result shows subdividing trade area with the proposed method is valuable for practical applications, and supports marketing decisions for Guiyang Tobacco Company.

Key Words: Marketing Decision, Micro Trade Area, Big Data Analytics, Subdividing

1. INTRODUCTION

The micro trade area (MTA) is a new kind of business model of electronic commerce. It uses APP technology of Wechat to help companies implementing their business on mobile terminals, and accomplishing network marketing. Conceptually, it comes from the traditional theory of trade area, but in practical sense, there are obviously differences between trade area and micro trade area, such as definition, usage mode. Thus the traditional partition method of trade area based on the geographic information theory could not be strictly applied to micro business trade. Based on the analysis of trade area theory, the division method of micro trade area was discussed. The trade area is a conception that has been examined in various contexts. Its original concept is defined as an urban area that contains the main concentration of commercial land use [13]. Several researchers describe it as an area marked by various qualitative indicators relative to the surrounding urban environment [14]. In recent year, it is described as unique area of massive concentration of activities and focus for the polarization of capital, economic and financial activities in cities [15]. Here we mainly discuss the relationship between retail stores and trade area, we take the definition of trade area as a part of a city or town's commercial, commercial fringe, and general commercial areas in which established to accommodate those retail businesses and services which are intended to serve requirements of a residential neighborhood, provide for visitor accommodations and tourist amenities, encourage pedestrian-oriented development and so on [3].This kind of trade area refers to a distribution of customers in an area that prefer to a certain store, which took as a center and extended along a certain direction and distance to cover a certain area, and called micro trade area. That means the micro trade area is also the geographical distribution of customers. Due to the sales activity of retail store usually has a certain geographical zone, it has a relatively stable micro district area [4]. And due to the different merchandise, traffic factors, geographical location and business scale, the size and form of micro trade areas varies greatly [5].Trade area analysis is an important topic of marketing campaigns, for cigarette marketing, it is very important in guiding the rationalization layout of tobacco retail store, providing the basis of strategy formulation of cigarette launch, and so on. In cigarette industry, the trade area of cigarette retailers is divided into 42 categories with town, rural area, large, medium and small scale 7 formats. Among them, the town and rural area are the dimensions relevant to classifications of trade area. Based on this classification, some of tobacco companies further subdivide the trade area into micro trade area such as government district, school district, commercial mall district, tourist attractions and soon. But in practical sense, the analysis results of this subdividing trade area are unable to satisfy the need of application. In many cases, the categories of these districts could not reflect the difference among retail stores in the same area, and its direct result is the same delivery strategy to be used in different retail stores in the same district, and could not obtain better practice in sales and marketing strategic planning for a company. For these reasons, this paper proposes a method to subdivide trade area based on big data analytics by focusing on the correlation and real-time data of trade area to improve the application value of trade area analysis. By quantifying the number of sales, sales amount and structure of cigarette retail store in same commercial facilities, and then categorizing the data according to the type of retail store and comparing data for the same trade area, it is possible to determine the impact of retail store type on marketing decisions.

2. LITERATURE REVIEW

The coming of Big Data age has become an indisputable fact [1]. The notion of "Big Data" is now being recognized broadly in the field of data-driven decision-making. There

978-1-4799-7016-2/15/$31.00 ©2015 IEEE

is growing enthusiasm for the value of data, and products obtained through analyzing it. According Gartner's definition, Big Data is high-volume, high-variety and high velocity information assets, there are many obvious problems impede progress at all phases of the pipeline that can create value from data, such as complexity, scale, and timeliness with Big Data [7]. Thus there are many challenges from transforming Big Data into a structured format to later analysis with these data. In many practical senses, due to lack of the complexity of the data that needs to be analyzed and scalability of the underlying algorithms, data analysis is a clear chokepoint. Here we only discuss the analysis methods with Big Data on micro trade area partition. Many methods are proposed to analyze data, such as A/B testing, association rule learning, cluster analysis, classification, and regression. However, these statistics and computer science (particularly machine learning) methods are tend to find the causal relationship among these data. In Big Data age, the data becomes more valuable when it can be linked with other data, thus data integration is a major creator of value. Thus the correlation is key point rather than causation among data. If users cannot understand the analysis, they can obtain limited value by analyzing Big Data. Ultimately, a decision-maker, provided with the not causation of analysis, has not ability to interpret these results. Usually, this involves examining all the assumptions made and retracing the analysis [8]. Finally, presentation of the results and its interpretation by non-technical domain experts is unable to extracting practical knowledge [9]. The A/B testing and regression analysis can be applied effectively to smaller datasets, but other techniques such as cluster analysis and classification can be strictly applied to analyze big data, and, in general, more numerous and insightful results can be generated by using larger and more diverse datasets than smaller, less diverse ones. In the field of trade area research, there are many classical works in trade area theory, such as Reilly and Covers' the laws of retail Gravitation [16], Huff's probability analysis of center trade area [17], The Index of Retail Saturation Theory [18] and so on. These works focus on the use of geographic information systems and spatial models in market area analysis [19][20].In the recent years, due to the pressure on city centers and the use of mobile intelligent terminals, activity and spaces of retail stores have changed drastically [11]. For any enterprise, knowing its retail store whether they could compensate for the requirements of changing expectations of the city dwellers or other people from trade areas (such as entertainment facilities, transportation stations, and industrial areas and so on) becomes a key point that effects the marketing policy [10].In summary, in the research field, surrounded the concepts of data, social impact, future trends, especially for social media developments and changes in the information age, Big Data makes the academic community to face the new opportunities and challenges [2]. Particularly, in the field of social sciences, Big Data approach will assist research in social science field. In the business field, using Big Data analysis approach to solve business problems has become a trend, such as consumption forecasts. Simply, the Big Data approach includes three aspects [6]:

1) Not a random sample, but all data, real-time data that are specific to individual samples analysis.

2) Not accurate, but mixed, particularly, the algorithm for large data more simple and effective than small data.

3) Not causation, but correlation.

The idea of subdividing the trade area to micro trade areas 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 mining.

Based on the three steps mention above, we use Big Data analysis for the MTA partition with the existed sales data, 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 store. It is obvious that making full use of these data can reflect the Big Data characteristics.

3. THE METHODOLOGY OF SUBDIVIDING TRADE AREA BASED ON BIG DATA ANALYTICS

In section 1, the traditional method for subdividing trade area cannot meet the practical needs have been discussed in the cigarette industry. From the perspective of specific implementation of Guiyang Tobacco Company, subdividing
the area into 20 categories is enough, but even that subdividing method still does not have separate data validation to display its significance for its retail stores. For this reason, this paper tentatively proposes a Big Data analysis method, which combined the data of a certain zone in the existing 20 trade areas with the cigarette sales data, amount of cigarettes, cigarettes structured data, and use the method of cluster analysis to subdivide each trade area for forming new sub trade areas, i.e., MTA. The method for subdividing trade area is as following with SPSS (Statistical Product and Service Solutions) in specific analysis process. The SPSS Statistics is a widely used software package for statistical analysis in social science, market researchers, marketing organizations, data miners, and others [12]. It can satisfy our need for analysis.

3.1 Data Preparation

A set of random sales data of cigarette retail stores in Guiyang city has been selected for analysis as in Table1 . We just list a part of the dataset. The data of retail stores are collected from V3 software system used in Guiyang Tobacco Company, containing 636 samples. Among the data items, the item No. labels uniquely a retail store. Area denotes trade area segments based on geographic location, and encoded by software for data analysis. Sales contains sales and sales amount. We take the data of Zone 1 as an example for introducing the MTA partition. For further discussion, we need to predispose the data as in Table 2 following next two steps:

1) Adding data: adding 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 and sales structure:

$$\text{Structure} = \text{Amount}/\text{Sales} \quad (1)$$

Based on formula (1), we could obtain the prepared three-month data of retail store for analysis.

2) Standardizing data: the measurement units of three data types, i.e., cigarette sales, sales amount and structure are different, thus numerical methods must be used to standardize all variables, i.e., dimensionless for solving the problem of values are not integrated. We use standard deviation to standardize the variables (z-score standardization) marked as Zscore.

Table2. Standardized Variables

No.	Area	Zscore (Average of sales)	Zscore (Average of sales amount)	Zscore (Average of sales structure)
1	Zone 1	3.30077	4.01877	1.96063
2	Zone 1	3.61170	3.89945	1.43629
3	Zone 1	2.99529	3.80155	2.26389
…	……	……	……	……

3.2 Using Clustering Algorithm to Subdivide Trade area

After the data is ready, we can use clustering algorithm to subdivide trade area.

1) Choosing subdivided trade area: taking Zone 1 as an example (contains 54 samples), subdividing trade area.

2) Selecting cluster variables and methods: taking cigarette sales, amount, structure as variables, using a hierarchical clustering method (also known as cluster algorithms) and K-Means clustering method (also called fast clustering) to analysis on the certain zone. The difference between the two methods is, the former need not preset the classification number, relatively complex, while the latter requires preset classification number, relatively simple.

In the first step, we select the hierarchical clustering to form a tree intuitively statement classification of the certain zone as shown in Figure 1. From figure 1 we can see: in the first layer of clustering, the zone is subdivided into two regions, No.335 retail store as a separate category, others retail stores as another category. In the second layer of clustering, the zone is subdivided into three, up to five layer and nine categories.

In the second step, based on hierarchical clustering, we use K-Means clustering method to subdivide the trade area, preset classification as three categories, and obtained eight retail stores in category 1, 45 retail stores in category 2 and 1 retail store in category 3. The number of various types of subdivided area as Table 3:

After the data is ready, we can use clustering algorithm to subdivide trade area.1) Choosing subdivided trade area: taking Zone 1 as an example (contains 54 samples), subdividing trade area.2) Selecting cluster variables and methods: taking cigarette sales, amount, structure as variables, using a hierarchical clustering method (also known as cluster algorithms) and K-Means clustering method (also called fast clustering algorithms) to analysis on the certain zone. The difference between the two methods is, the former need not preset the classification number, relatively complex, while the latter requires preset classification number, relatively simple. In the first step, we select the hierarchical clustering to form a tree intuitively statement classification of the certain zone as shown in Figure 1. From figure 1 we can see: in the first layer of clustering, the zone is subdivided into two regions, No. 335 retail store as a separate category, others retail stores as another category. In the second layer of clustering, the zone is subdivided into three, up to five layers and nine categories. In the second step, based on hierarchical clustering, we use K-Means clustering method to subdivide the trade area, preset classification as three categories, and obtain eight retail stores in category 1, 45 retail stores in category 2 and 1retail store in category 3. The number of various types of subdivided area is listed as Table 3:

Table3. The Classification Number of Subdivided Trade Area

The number of cases in each cluster		
Cluster	1	8
	2	45
	3	1
Effective Cases	54	
Dismissed Cases	0	

In summary, the results obtained from hierarchical clustering method and K-Means clustering method is consistent. Compared with traditional subdividing 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) Subdividing the trade area more detailed, for example, subdividing Zone 1 , hierarchical clustering method can subdivide the 54 retail stores in Zone 1 up to into 9categories, even meticulous to subdivide a retail store as a category of trade area (for example, retail store No. 335). Of course, in the application of subdividing trade area classification, it needs to consider the value that subdividing a retail store as a classification of trade area in a practicalsense.3) The proposed method for subdividing the trade area 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 subdivision method. In summary, the 54 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. By subdividing the district, the MTA partition can support the manufacturers or retailers adopting more marketing strategies to serve the customers in a traditional trade area.

3.3 The Validation of the MTA Partition Method

To ensure the consistency of the results, we use curve fitting method to verify the goodness of fit of the MTA

partition method. Using SPSS, we obtain the linear regression result as in Table 4, 5 and 6.

Table4. The Model Summary of Linear Regression Test on the MTA Partition

Model Summary				
Model	R	R^2	Adj R^2	Standard error of estimate
1	.906[a]	.821	.810	.170

a. Predictors: Constant, Zscore (Average of sales structure), Zscore (Average of sales), Zscore (Average of sales amount).

Table5. 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	6.642	3	2.214	76.3	.000[a]
Residuals	1.450	50	.029		
Total	8.093	53			

a. Predictors: Constant, Zscore (Average of sales structure), Zscore (Average of sales), Zscore (Average of sales amount).

b. Dependent variable: Category No. of cases.

c. Model: 1

Table6. The Coefficient of Linear Regression Test on the MTA Partition with Model 1

Coefficient					
Model	Non standardized coefficients		Standardized coefficients	t	Sig.
	β	Standard error	Trial version		
Constant	1.82	.024		77.1	.000
Zscore (Average of sales)	.743	.130	1.548	5.73	.000
Zscore (Average of sales amount)	-1.3	.146	-2.501	-8.8	.000
Zscore (Average of sales structure)	.171	.029	.510	5.9	.000

Dependent variable: Category No. of cases.
Model: 1

From Table 4 to Table 5, we note that:

1) R^2=0.821, it points out that 82.1% of the variation of 3 categories in Zone 1 can be explained by three sub-indexes, or the three sub-indicators can explain 82.1% of the variation of subdivided trade area.

2) The values of Sig are low. It indicates that assumptions are significant, i.e., by subdividing the trade area, the differences between the different trade areas are more obvious, that means the difference presents different cigarette sales. From the Table 6, the regression equation can be obtained as following:

$$\ln Y = 1.82 + 0.74 \ln X_1 - 1.3 \ln X_2 + 0.17 X_3 \quad (2)$$

In the formula (2), $\ln Y$ represents the classification, X_1 represents cigarette sales, X_2 is on behalf of cigarette sales amount, X_3 represents the cigarette sales structure. Among them, the sales and amount of cigarette sales impact on the classification are 0.74 and -1.27 (negative correlation), greater than the impact of cigarette structure (0.17), then, in practical sense, the impact of cigarette structure can be ignored. The characteristics of the three subdivided trade area can be determined:

Category 1: the amount of sales is relatively high, relatively few sales, these retail stores are belong to high-end trade area.

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

Category 3: a relatively small sales and amount of sales, these stores belong to low-end trade area.

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

4. Conclusion

In this paper, based on the traditional method of subdividing trade area, we take the practical sales data of Guiyang Tobacco Company as an example, and propose a method called the MTA partition with clustering algorithm for subdividing trade area using Big Data analysis. By validating the proposed method, we conclude 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 subdividing trade area methods, the method has more significant advantages in determining retail stores categories, prediction on cigarette sales and structure, cigarette delivery decision and other aspects 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 more independent variables in subdividing trade area, the proposed method will certainly have further application effects.

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, Vol.50, No.1, 146-169, 2013.

[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, Vol.9, No.6, 317-326,2002.

[4] M. Kakoi, S. Saito, Extracting rules of loyal visitors' characteristics at central trade areas based on

on-sitesampling, Studies in Regional Science, Vol.36, No.4,993-1005, 2006.

[5] M. Angélica Salazar-Aguilar, Roger Z. Ríos-Mercado and José L. González-Velarde, A bi-objective programming model for designing compact and balanced territories in trading area, Transportation Research Part C: Emerging Technologies, Vol.19, No.5, 885-895, 2011.

[6] N. Marz, J. Warren, Big Data: Principles and Best Practices of Scalable Real-time Data Systems, Manning Publications, New York, America, 2012.

[7] J. Manyika, M. Chui, Big Data: The next frontier for innovation, competition and productivity, McKinsey Quarterly, No.5, 27-30, 2011.

[8] N. Archak, A. Ghose, Deriving the pricing power of product features by mining consumer reviews, Management Science,Vol.57, No.8, 1485-1509, 2011.

[9] J. Howe, Crowd sourcing: How the Power of the Crowd Is Driving the Future of Business, Crown Publishing Group, New York, America, 2006.

[10] S. E. Kilic, Preservation plan applications for the historical city centre, Kemeralti (Izmir, Turkey), European Planning Studies, Vol.16, No.2, 253-276, 2008.

[11] C. J. Lopes Balsas, City center revitalization in Portugal, Cities, Vol.17, No.1, 19-31, 2000.

[12] J. Pallant, SPSS Survival Manual, Open University Press, New York, America, 2011.

[13] R. W. McColl, Encyclopedia of World Geography, Infobase Publishing, New York, America 2005.

[14] L. Halbert, The Paris Region: Polycentric Spatial Planning in Monocentric Metropolitan Region. In P. Hall and K. Pain (Eds.), The Polycentric Metropolis: Learning from Mega-city Regions in Europe. Earthscan Publications, London, British, 2006.

[15] M. Drozdz, M. Appert, Re-understanding the CBD: A Landscape Perspective. In D. Naik and T. Oldfield (Eds.), Critical Cities. Myrdle Court Press, London, British, 2010.

[16] P. D. Covers, New laws of retail gravitation, Journal of Marketing, Vol.14, No.3, 379-384, 1949.

[17] D. L. Huff, Parameter estimation in the Huff Model, ArcUser, October-December, 34-36, 2003.

[18] Z. W. Chen, Retail Marketing: From Practical Views, Beijing University Press, Beijing, China, 2008.

[19] G. Dudley, Probability Primer: Spatial insights into StoreBase Trading Areas, A Publication of Blackburn/Polk Marketing Services Inc., Toronto, America, 1996.

[20] S. Fotheringham, M. O'Kelly, Spatial Interaction Models: Formulations and Applications, Kluwer Academic Publishers, Dordrecht, Holland, 1989.

Table1. Sales Data of Retail Stores from January to March in 2014

No.	Area	January sales (per carton)	January sales amount (yuan)	February sales (per carton)	February amount (yuan)	March sales (per carton)	March sales (yuan)
1	Zone 1	475	70222.00	335	50372.00	325	52025.00
2	Zone 1	495	67139.50	340	45688.00	455	52280.50
3	Zone 1	805	99112.00	480	53074.00	370	35302.00
4	Zone 1	250	26822.00	130	13366.00	140	13709.00
5	Zone 1	2090	336090.00	500	56489.00	620	70280.00
6	Zone 1	435	47477.00	190	17802.00	195	16847.00
7	Zone 1	895	99968.00	125	13534.00	600	68007.00
8	Zone 1	260	29342.50	85	9436.50	165	15520.50
9	Zone 1	1305	112004.00	460	32063.00	390	20111.00
10	Zone 1	600	81689.00	100	12344.00	280	31833.00
……	……	……	……	……	……	……	……

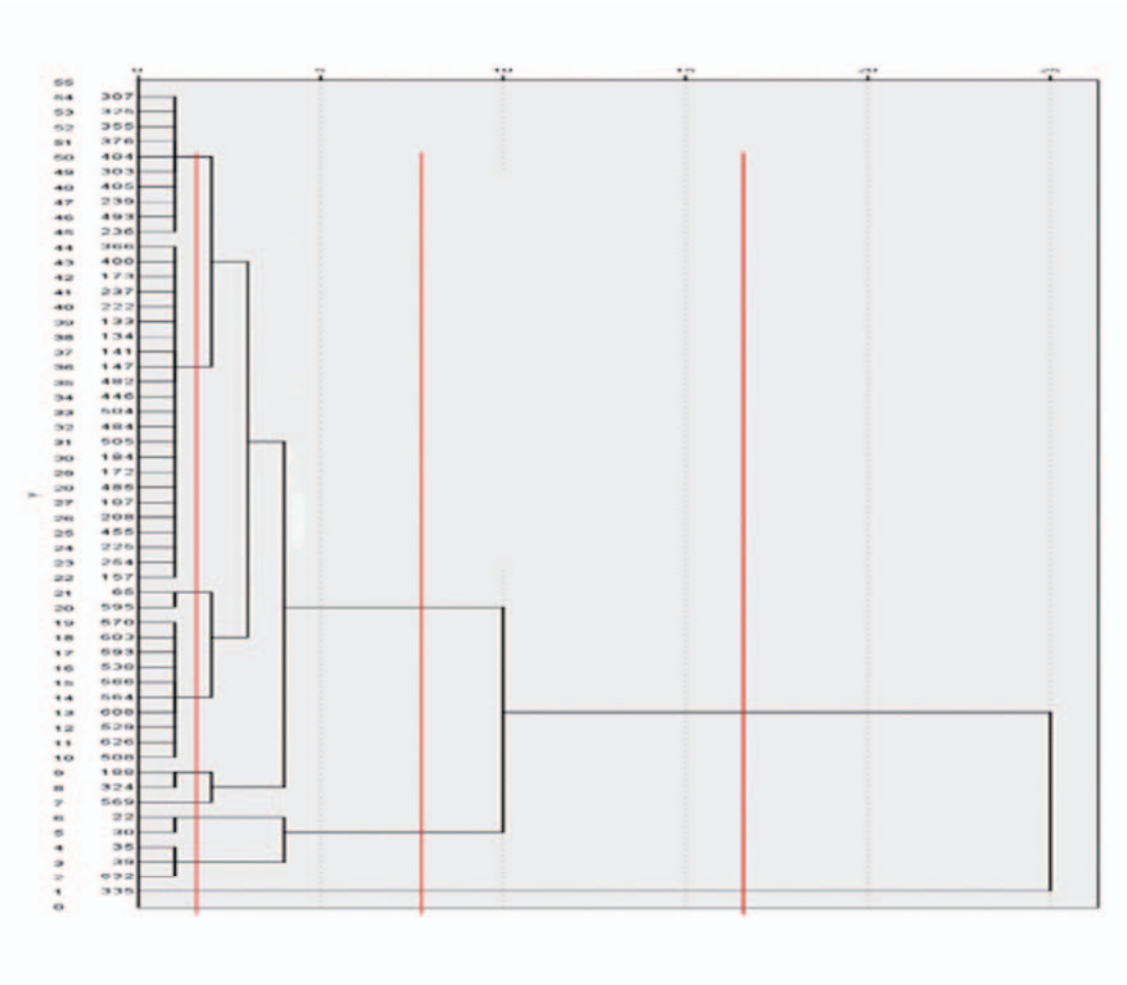

Fig.1: The hierarchical clustering tree of Zone 1