International Journal of Smart Business and Technology
Vol. 5, No.2(2017), pp. 21-26
http://dx.doi.org/10.21742/ijsbt.2017.5.2.04

A Study on the Improvement of Algorithm Using Beacon Data in Consumer Moving Path Similarity Measurement

MyeongBae Kim[1], Sung Won Jung[2], Yong Deuk Ji[3], Jong Bae Kim[4]and Gwang Yong Gim[5]

[123]06978 Dept. IT Policy Management, SoongsilUniv, Sangdo-dong, Dongjak-Gu, Seoul, Korea
[4]06978 Dept. Graduate School of Software, Soongsil Univ., Sangdo-dong, Dongjak-Gu, Seoul, Korea
[5]06978 Dept. Business Administration, SoongsilUniv, Sangdo-dong, Dongjak-Gu, Seoul, Korea
[1]mbkim77@gmail.com, [2]jdate@naver.com, [3]ofcyd@daum.net, [4]kjb123@ssu.ac.kr, [5]gygim@ssu.ac.kr

Abstract

This study suggested an improved CMPSI(Consumer Moving Path Similarity Index) algorithm to apply the characteristics of online shopping malls and beacon data. This algorithm is a standardized index to define standardized similarity considering the route in the offline area and detention time in each beacon zone and to combine with similarity considering the definition. The method of similarity measurement suggested by this study might be expected to utilize beacon service, to be used as important information for marketing activity to increase revenue through combining with on/offline purchase information, and to be applied to other major offline B2C commercial area.

Keywords*: B2C commercial area, beacon data, Consumer Moving Path Similarity Index*

1. Introduction

From the perspective of a marketer analyzing a customer, they will want to know the characteristics of their customers. Beacon technology is one of the technologies that can collect such information.

Thus this study is going to verify the practicality of this method by suggesting the similarity algorithm considering the travel route and the time required for practical marketing activities in large-scale offline stores and applying it to actual data. In addition, this study is going to suggest a method for analyzing the data collected by BLE Beacon for effective marketing activities in the offline store, and propose the fields that can be extended.

2. Related researches on similarity algorithm

Article history:
Received (November 2, 2017), Review Result (November 12, 2017), Accepted (November 22, 2017)

ISSN: 2288-8969, eISSN: 2207-516X IJSBT

Hirschberg, D. S. (1977) suggested LCSS (Longest Common Subsequence) function to measure the similarity between character strings. This method finds the longest sequence among the partial sequences common to both strings of sequences [1].

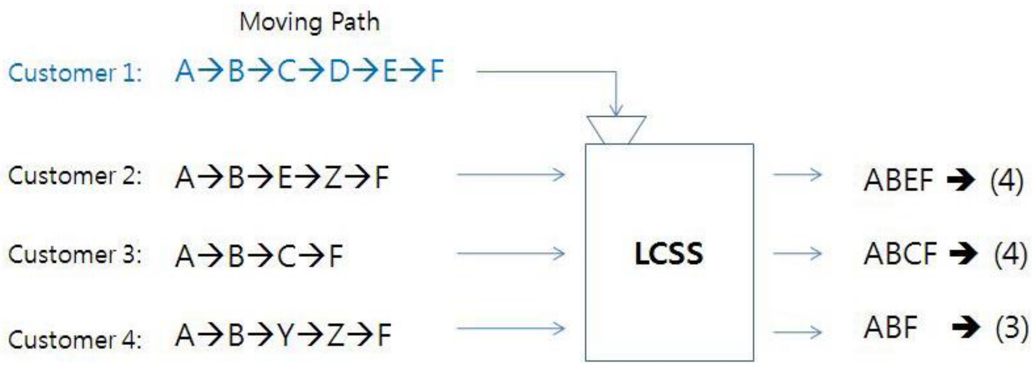


Figure 1. The measure example of LCSS

When we apply this method to the customer's moving path in the offline space, a couple of problems arise. It is that it is highly likely that the LCSS length will be large among the individuals having long travel length. To overcome these shortcomings, Jung In-cheol (2012) suggested relative_LCSS that standardized LCSS values [2].

It is to correct the total length of travel that the two objects moved to the length of LCSS length. The similarity function of relative_lcss that changed LCSS is as follows.

Here, length(x), length(y) are the travel distance of each x and y.

$$relative\_lcss(x,y) = \frac{lcss(x,y)}{length(x)+length(y)} \qquad \text{[Formula 1]}$$

$$CVTI(a,b) = \sum_{i=1,j=1}^{n,m} |a_i.I \cap b_j.I| \quad ,where\ a_i.c = b_i.c \qquad \text{[Formula 2]}$$

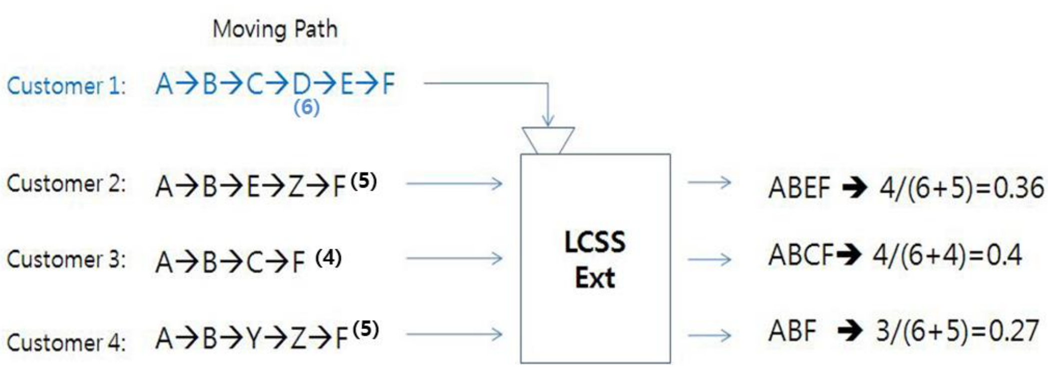
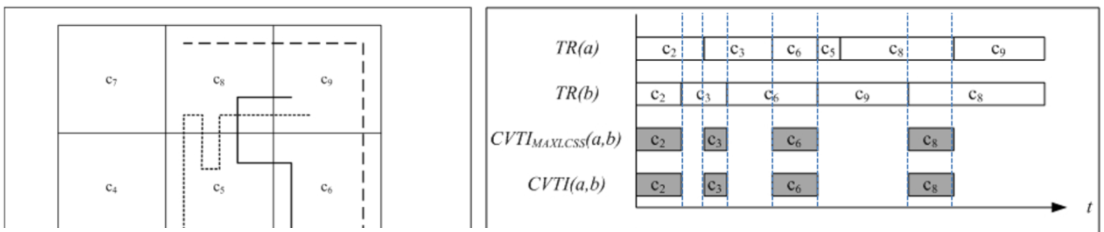


Figure 2. The measure example of relative_LCSS

$$S_{LCSS(x,y)} = c_{ij} = \begin{cases} 0 & \text{if } i=j=0 \\ c_{i-1,j-1} & \text{if } i,j>0 \text{ and } x_i = y_j \\ MAX\{c_{i-1,j}, c_{i,j-1}\} & \text{if } i,j>0 \text{ and } x_i \neq y_i \end{cases} \qquad \text{[Formula 3]}$$

$$Std-relative\_S_{LCSS} = \frac{S_{LCSS(x,y)}}{Min(length(x),\ length(y))} \qquad \text{[Formula 4]}$$

Kang Hyeyoung et al. (2008) suggested a similarity algorithm considering the Common Visit Time Interval (CVTI) under the cell-space basis. [3].



$$adj-CVIT(x,y) = \left\{ \sum_{i=1,j=1}^{n,m} |x_i.I \cap y_j.I| \right\} / MIN(T_x, T_y) \qquad \text{[Formula 5]}$$

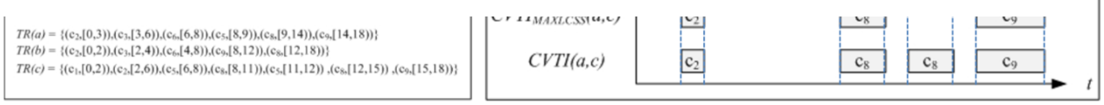


Figure 3. The measure example of CVTI

3. Improvement of moving path similarity algorithm

3.1. Improvement of moving path similarity algorithm considering travel route

The existing relative_LCSS does not have a value between 0 and 1 exactly. Thus, this study use MIN(length(x),length(y)) in denominator for correct standardization. That is, the minimum value of the number of visited paths of two objects is used.

Here, to standardized-relative_LCSS with the value between 0 and 1, when using the minimum value in the denominator, it can be expressed as follows.

3.2. Improvement of moving path similarity algorithm considering stay time

The stay time means the time that stayed in the zone where the beacon is installed. The beacon zone is an area configured with a predetermined radius and records the time at which the customer's mobile enters and the time the customer departs the area. Departure Time - Entry time is defined as Stay Time. The stay time similarity is defined in consideration of the stay time of the zone included in LCSS(x,y), regardless of the elapsed time from the initial entry time of the shopping mall based on the characteristics of shopping. This can be expressed as follows.

3.3. Proposed model of mixed similarity algorithm

This study proposes a customer moving path similarity measurement algorithm for a large off-line shopping mall by combining standardized-relative_LCSS and adj_CVTI suggested in Chapter 3. Both values have a value between 0 and 1. Therefore, this study proposes a Consumer Moving Path Similarity Index (CMPSI) algorithm with a linear combination of two successive similarities. Here, if the sum of the weights of two arithmetic formulas is set to 1, the CMPSI is also an index having a value between 0 and 1. This can be expressed as [Formula 6]. Both values have a value between 0 and 1. Therefore, this study proposes a Consumer Moving Path Similarity Index (CMPSI) algorithm with a linear combination of two successive similarities. Here, if the sum of the weights of two arithmetic formulas is set to 1, the CMPSI is also an index having a value between 0 and 1. This can be expressed as follows.

Where, $S_{LCSS(x,y)}$: The length of the longest common substring of object x and y

$Min(length(x),\ length(y))$: Minimum value of the number of the travel routes of x and y

$\sum_{i=1,j=1} |x_i.I \cap y_j.I|$: Common Stay time of the Same Zone of a and b objects

$MIN(T_x, T_y)$: Minimum Value of Total Shopping Time of Object x and y

α, β : As for weighted value of each α+β=1, 0 ≤α, β≤ 1

4. Empirical analysis

[Table 1] shows the characteristics of males who did not show characteristics by age group. For males, the moving path similarity of groups with the same sex and age group is high for all ages. On the other hand, in the case of women, the same sex and age groups have high similarity only in some the 40s and 20s, and the rest have high similarity among the sex/age groups. This could be the result deducted because the patterns of shopping are simple for men but they are very diverse for women.

These results show that the newly proposed algorithm is worthy of exploiting the moving path characteristics of customers.

$$CMPSI = \alpha\,(Std-relative\_S_{LCSS}) + \beta\,(adj-CVIT(x,y))$$

$$= \alpha \frac{S_{LCSS(x,y)}}{Min(length(x),\ length(y))} + \beta \left\{ \sum_{i=1,j=1}^{n,m} |x_i.I \cap y_j.I| \right\} / MIN(T_x, T_y) \qquad \text{[Formula 6]}$$

Table 1. Similarity comparison by gender, age

Gender	Age	Gender/Age Same	adj_lcss average	adj_CVTI average	CMPSI average	Number of comparison customers
Male	20s	Same	0.341	0.239	0.290	10
		Different	0.281	0.222	0.251	233
	30s	Same	0.277	0.219	0.248	152
		Different	0.261	0.228	0.245	853
	40s	Same	0.352	0.276	0.314	9
		Different	0.334	0.265	0.299	352
	50s	Same	0.500	0.317	0.409	1
		Different	0.385	0.308	0.346	76
Female	20s	Same	0.287	0.226	0.257	100
		Different	0.276	0.238	0.257	721
	30s	Same	0.248	0.229	0.238	529
		Different	0.289	0.261	0.275	1,473
	40s	Same	0.334	0.280	0.307	274
		Different	0.286	0.245	0.265	1,135
	50s	Same	0.266	0.194	0.230	44
		Different	0.280	0.246	0.263	279

5. Conclusion

This study suggested a new algorithm to calculate the similarity based on the time of staying at a specific location and the moving path in a large offline shopping mall. This study suggested an improved algorithm after pointing out the inappropriate part of the similarity algorithm suggested in existing studies, from the viewpoint of the customer moving path of the large offline shopping mall. Also, this study applied to the beacon data using the data occurring from the beacon, and as a result, it could be utilized as a customer identification index that can be used for marketing.

In this study, if we identify the moving path of the visiting customers in the large commercial space and find the characteristic of each zone based on this, and apply appropriate marketing to the characteristic, it will be possible for customers to expect a more convenient space and for companies to expect profit increase.

References

[1] Hirschberg, D. S. "Algorithms for the longest common subsequence problem". Journal of ACM, Vol.24, No.4, pp.664-675 **(1977)**

[2] J. Incheol. "A Study on moving path Data Clustering Techniques for Moving Objects: Case Study: Moving path Analysis of Large Distribution Store Customers". Dongguk University Doctoral Thesis **(2012)**

[3] Kang H., Kim J., H. Jung Rae, K.J. Lee. "Similarity Measure of Moving Object Trajectories in Cellular Space". Journal of Korea Geographic Information Systems, Vol. 16, No. 3, pp.291-301**(2008)**

[4] K. Daewon. "A study on indoor location based dementia patient monitoring system using Bluetooth beacon". Digital Convergence Research, Vol. 14, No. 2, pp.217-225 **(2016)**

[5] Kim S., Park G., Koo Hy., Kim S. "An Exploratory Case Study on the Improvement of Omni Channel Service Type and Shopping Experience - Focusing on the Three Domestic Major Distribution Centers" (Lotte, Shinsegae, Homeplus). Design Convergence Research, Vol. 15, No. 5, pp.85-103 **(2016)**

[6] Y. Seungjoon, J. Inchul, and K. Youngshik. "Customer shopping moving path pattern analysis using RFID data". Journal of the Korean Society for Information Science and Technology, Vol.11, pp.61-74 **(2012)**

[7] Y. Minseon. "Understanding Consumer Showrooming Behavior Applying Rational Behavior Theory. Distribution Studies", Vol. 20, No. 4, pp. 79-103, **(2015)**

[8] O. Amseok. "Smart Factory Logistics Management System Using Bluetooth Beacon Based Indoor Location Tracking Technology". The Journal of the Korea Information and Communications Society, Vol. 19. **(2015)**

[9] C. Hyunseung, Y. Sungbyeong. "A Study on the Factors Affecting the Intention of Switching from Online Shopping to Webrooming". Intelligent Information Research, Vol. 22, No. 1, pp.19-41, **(2016)**

[10] Cao, H., Mamoulis, N., & Cheung, D. W. "Discovery of periodic patterns in spatiotemporal sequences". IEEE Transactions on Knowledge and Data Engineering, Vol.19, No.4, pp.453-467, **(2007)**

[11] Larson, J. S., Bradlow, E. T., & Fader, P. S. "An exploratory look at supermarket shopping paths". International Journal of research in Marketing, Vol.22, No.4, pp.395-414, **(2005)**

[12] Tung, A. K., Hou, J., & Han, J. "Spatial clustering in the presence of obstacles". In Data Engineering, Proceedings. 17th International Conference, on IEEE pp. 359-367, **(2001)**