

# Finding the Best Location for Logistics Hub Based on Actual Parcel Delivery Data

Ha Yoon Song(✉) and Insoo Han

Hongik University, Mapo-gu, Seoul, Republic of Korea
hayoon@hongik.ac.kr, haninsoo9989@gmail.com

**Abstract.** So many national and international packets are traveling around in this time. The parcel delivery service is a major part of nationwide logistics. It is reported that wrong routes for logistic causes economical disadvantage both in time and in cost. It is possible to collect actual delivery data from logistics company or Internet of Things devices. Based on actual route of packet delivery, we collected 100,000 delivery data over Republic of Korea and analyzed for optimal hub candidate locations in terms of minimum distance and minimum time. From the raw delivery data set, actual delivery paths were calculated in terms of big data analytics Using Longest Common Route Subsequence algorithm, the most common paths can be identified. From the economic aspect, regarding minimum distance and time, optimal hub location candidates were voted and identified. With several hub locations, optimal distance and time can be calculated from the location of optimal hub candidates.

**Keywords:** Parcel service · Optimum logistics hub location · Longest Common Route Subsequence algorithm · Big data analytics

## 1 Introduction

Parcel delivery service is one of major axis in national or international economy. Investigating lots of delivery route shows that improper routes of parcel delivery rises costs extra distance and time for parcel delivery and causes economical disadvantage of nations. The hub and spoke method of parcel delivery is commonly used in most of nations [1,2]. It does mean, the location of hub is one of the critical parameter for economical parcel delivery [3]. In this paper, we tried to determine an optimum location of logistics hub based on actual delivery data. From the parcel delivery company, we collected 100,000 delivery data including source and destination, actual routs of delivery, current location hub and others. The data is collected in Republic of Korea and only public data was utilized. The area size of Republic of Korea is 1,002,000 ha and 5,100,0000 population

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (NRF-2019R1F1A1056123).

© Springer Nature Switzerland AG 2019
S. Misra et al. (Eds.): ICCSA 2019, LNCS 11619, pp. 603–615, 2019.
https://doi.org/10.1007/978-3-030-24289-3_45

which leads to high density population. The major purpose of this research is to avoid inefficient delivery cost by selecting a good hub location with minimum distance and minimum time for delivery. Very few research dealt with this topic. For example, Yunhe Ma of Beijing Jiaotong University showed an mathematical approach over virtual data set in order to figure out best hub locations [4]. Other researches done at National Taiwan University by Yon-Chun Chou analyzed hub transshipment across flexible time periods in order to decrease workload and sojourn time [5].

This paper is structured as follows. Section 2 discuss about the process outline of this research including algorithm we used. Section 3 visualizes results on actual maps, according to the process shown in Sect. 2. Section 4 shows logistics hub candidates based on the previous analysis results, as well as benefits in terms of time and distance. We will conclude this paper in Sect. 4 with future direction of further research.

## 2 Process Outline

We collected 100,000 raw data of parcel delivery. Then the actual route of parcel delivery will be collected. Based on the Longest Common Route Subsequence (LCRS) algorithm, we will deduce the most populated locations of parcel delivery. Among the most populated locations, several locations can be identified by voting mechanism. Then the most populated locations be reduced to candidates of optimal logistics hub. From the concepts of Longest Common Sequence (LCS) by Bergroth [6] and Sebastian [7], LCRS algorithm were developed for this research.

### 2.1 Source Data Set

Table 1 shows sample data of parcel delivery route. Pathnum stand for parcel identifier, category stands for parcel contents category. The column p_order

**Table 1.** Sample source data

| Pathnum | Category | p_order | Name | Time | lat | lng |
|---|---|---|---|---|---|---|
| 620703310075 | Digital | 0 | start point | 201901171735 | 35.1002578 | 129.0264814 |
| 620703310075 | Digital | 1 | waypoint1 | 201901180244 | 36.2518725 | 127.6236381 |
| 620703310075 | Digital | 2 | waypoint2 | 201901181056 | 38.2070148 | 128.5918489 |
| 620703310075 | Digital | 3 | waypoint3 | 201901181346 | 38.1673391 | 128.516669 |
| 620703310075 | Digital | 4 | end point | 201901181648 | 38.3801292 | 128.4674385 |
| ... | ... | ... | ... | ... | ... | ... |
| 620703311442 | Apparel | 0 | start point | 201901170745 | 35.1002578 | 129.0264814 |
| 620703311442 | Apparel | 1 | waypoint1 | 201901180218 | 36.2518725 | 127.6236381 |
| 620703311442 | Apparel | 2 | waypoint2 | 201901180823 | 35.1761938 | 129.0797244 |
| 620703311442 | Apparel | 3 | waypoint3 | 201901180838 | 35.1088179 | 129.0101733 |
| 620703311442 | Apparel | 4 | end point | 201901181044 | 35.1088179 | 129.01017335 |

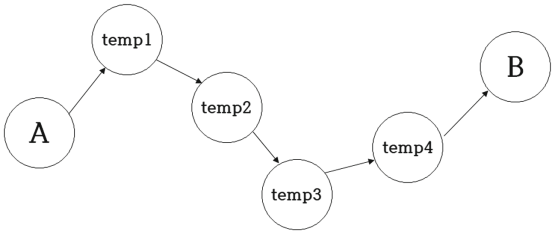


**Fig. 1.** Example of route with subsequences

stands for specific route point, such that 0 for starting point, 1 for waypoint 1, 2 for waypoint 2, 3 for waypoint 3 and 4 for destination of parcel routes. If p_order contains 4, it does mean the parcel has been delivered. The column name represents actual location name. Time is for time at the location. Latitude is abbreviated as lat and longitude is abbreviated as lng. By preprocessing of 100,000 parcel deliver route data, 20,000 data is found valid for this research.

### 2.2 Route Data Set

One whole parcel delivery route tends to be divided into several sub routes as shown in Fig. 1. Table 2 shows sample route data. The column pnum shows the

**Table 2.** Sample route data

| Pathnum | pnum | Type | s_address | e_address | Duration | dist | num | Route |
|---|---|---|---|---|---|---|---|---|
| 620703310576 | 1 | 2 | start point | waypoint1 | 1:40 | 14 km | 0 | start point |
| 620703310576 | 1 | 2 | start point | waypoint1 | 1:40 | 14 km | 1 | temp1 |
| 620703310576 | 1 | 2 | start point | waypoint1 | 1:40 | 14 km | 2 | temp2 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 620703310576 | 1 | 2 | start point | waypoint1 | 1:40 | 14 km | 30 | temp30 |
| 620703310576 | 2 | 2 | waypoint1 | waypoint2 | 2:38 | 23 km | 0 | temp31 |
| 620703310576 | 2 | 2 | waypoint1 | waypoint2 | 2:38 | 23 km | 1 | temp32 |
| 620703310576 | 2 | 2 | waypoint1 | waypoint2 | 2:38 | 23 km | 2 | temp33 |
| 620703310576 | 2 | 2 | waypoint1 | waypoint2 | 2:38 | 23 km | 3 | temp34 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 620703310576 | 2 | 2 | waypoint1 | waypoint2 | 2:38 | 23 km | 28 | temp59 |
| 620703310576 | 3 | 2 | waypoint2 | waypoint3 | 1:20 | 12 km | 0 | temp60 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 620703310576 | 3 | 2 | waypoint2 | waypoint3 | 1:20 | 12 km | 14 | temp73 |
| 620703310576 | 4 | 2 | waypoint3 | end point | 0:48 | 8 km | 0 | temp74 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 620703310576 | 4 | 2 | waypoint3 | end point | 0:48 | 8 km | 6 | temp80 |
| 620703310576 | 5 | 1 | start point | end point | 5:10 | 48 km | 0 | start point |
| 620703310576 | 5 | 1 | start point | end point | 5:10 | 48 km | 1 | temp1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 620703310576 | 5 | 1 | start point | end point | 5:10 | 48 km | 23 | end |

sequence number of each subroute. For example, pnum 1 stands that the parcel is on the subroute from start point to waypoint 1, and pnum 3 stands at the parcel is on the subroute from waypoint 2 to waypoint 3 and so on. The exceptional pnum 5 stands for direct route from start point to end point, which is type 1. Once there exists waypoint in the whole route, the type is 2. The column num stands for subroute number. Routs column includes actual address of the corresponding location.

**Algorithm 1.** Longest Common Route Subsequence Algorithm

**Input:** route sequence for pathnum A and pathnum B
**Output:** A Set of Longest common Route subsequence, A Set of position
**PreProcessing(**$Path$**)** {
dp[len($A$)][len($B$)] // two dimentional array to check if two route sequence are identical
lcrs[len($A$)][len($B$)] // two dimentional array for common route subsequence
**for** (p=0; p < len($A$); p++) **do** // len($A$) = length of array A
    **for** (q=0; q < len($B$); q++) **do**
        **if** A[p]=B[q] **then**
            dp[p+1][q+1] = dp[p][q]+1
            **if** A[p] != Null **then**
                lcrs[p+1][q+1] = lcrs[p][q]+A[p]
            **end if**
        **else**
            dp[p+1][q+1] = maximum of $dp[p][q+1]$ and $dp[p+1][q]$
            lcrs[p+1][q+1] = lcs[p][q+1] or lcrs[p+1][q]
        **end if**
    **end for**
**end for**
**return** [lcrs]
}

### 2.3 Longest Common Route Subsequence

We developed LCRS algorithm as shown in Algorithm 1. The result of LCRS algorithm on the route data set is shown in Table 3. The time complexity of algorithm is $O(n^2)$ for the number of data, $n$. The column type is the same to that of Table 2. The column index stands for the sequence index of common subsequence between pathnum_A and pathnum_B. For example, for two parcels with pathnum_A and pathnum_B, index of 11 stands for 11 common location in the longest common route subsequence between pathnum_A and pathnum_B.

**Table 3.** Sample LCRS data

| pathnum_A | pathnum_B | Type | Index | Route |
|---|---|---|---|---|
| 620703310576 | 620703310580 | 1 | 1 | 48 Gumdannam-ro, Hanam-si |
| 620703310576 | 620703310580 | 1 | 2 | Geomdan Mountain Road |
| 620703310576 | 620703310580 | 1 | 3 | Dong Seoul Tolgate |
| 620703310576 | 620703310580 | 1 | 4 | Deokgye Temple Street |
| ... | ... | ... | ... | ... |
| 620703310576 | 620703310580 | 1 | 11 | 92 Deokgye-ro, Yangsan-si |
| 620703310576 | 620703310580 | 2 | 1 | 48 Gumdannam-ro, Hanam-si |
| 620703310576 | 620703310580 | 2 | 2 | Geomdan Mountain Road |
| ... | ... | ... | ... | ... |
| 620703310576 | 620703310580 | 2 | 5 | Central Road |

### 2.4 Voting for Optimal Hub Location

The voting is done by the count of location visit for parcel delivery. Table 4 shows sample voting results. The columns voting_range_start and voting_range_end stands for range of parcel identification number for voting. Table 4 shows that 100,000 parcels in a range of [620703310000, 620703410000] were searched for voting. As we stated already, 20,000 data out of 100,000 were utilized in our whole process. The column location means actual location in a path. The column counts the number of votes, i.e. higher count stands for higher population of parcel in the corresponding location. Algorithm 2 is an algorithm for this voting mechanism. It works on SQL of LCRS database and results in voting count of locations.

**Table 4.** Sample voting data

| voting_range_start | voting_range_start | Location | Count |
|---|---|---|---|
| 620703310000 | 620703410000 | Gyeongbu Expressway | 86434 |
| 620703310000 | 620703410000 | Yeongdong Expressway | 57717 |
| 620703310000 | 620703410000 | Gunjin 2-gil | 53415 |
| ... | ... | ... | ... |
| 620703310000 | 620703410000 | Jamsil Station | 11 |

**Algorithm 2.** Voting

**Input:** A Set of Longest common Route subsequence
**Output:** A Set of Voting
**PreProcessing(***voting***)** {

```
inputDB.execute("SELECT route FROM position")
// inputDB is database of input data
route = inputDB.fetchall() // route is route data array of input data
for (i=0; i < len(route); i++) do // len(A) = length of array A
    outputDB.execute("SELECT count FROM position WHERE route = 'route[i]'") // outputDB is database of output data
    routeCount = outputDB.fetchall()
    if routeCount == 0 then
        values = [(route[i], 1)]
        outputDB.execute("INSERT INTO position values (?,?)", values)
        outputDB.commit()
    else
        outputDB.execute("UPDATE position SET count = 'count' WHERE route = 'route[i]'")
        outputDB.fetchall()
    end if
end for
return [Voting]
}
```

## 3 Experimental Results

In this section, as we defined process outline in Sect. 2, we will provide results in a visual manner for each process steps. For the visualization, only part of data is utilized in order to have more clear figures.

### 3.1 Visualization of Source Data

Figure 2 shows mapping of raw parcel route data, and it is easy to figure out highly populated routes. There is one central location of highest population and there is actually a logistics hub named Okcheon hub. In case that there is another hub, the total costs for logistics could be decreased. Our purpose is to find location candidate for another hub as we will discuss in later subsections.

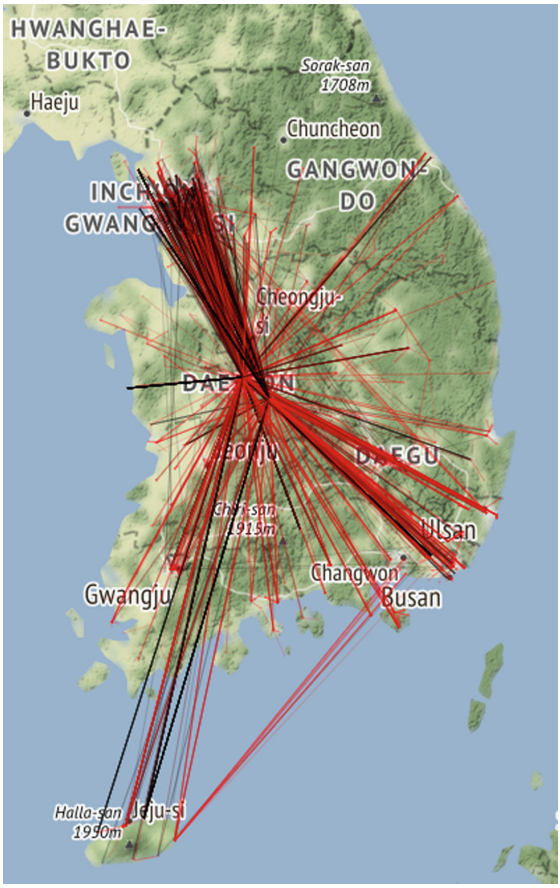


**Fig. 2.** Map with source data

### 3.2 Visualization of Route Data

Table 1 shows the density of route population with 10% of source data. The darker area shows the higher population of parcel route. The darkest area shown in the center of Fig. 3 locates Okcheon hub. Another darkest area in upper left corner locates capital city of Republic of Korea, Seoul. Lower right area with higher population locates Busan.

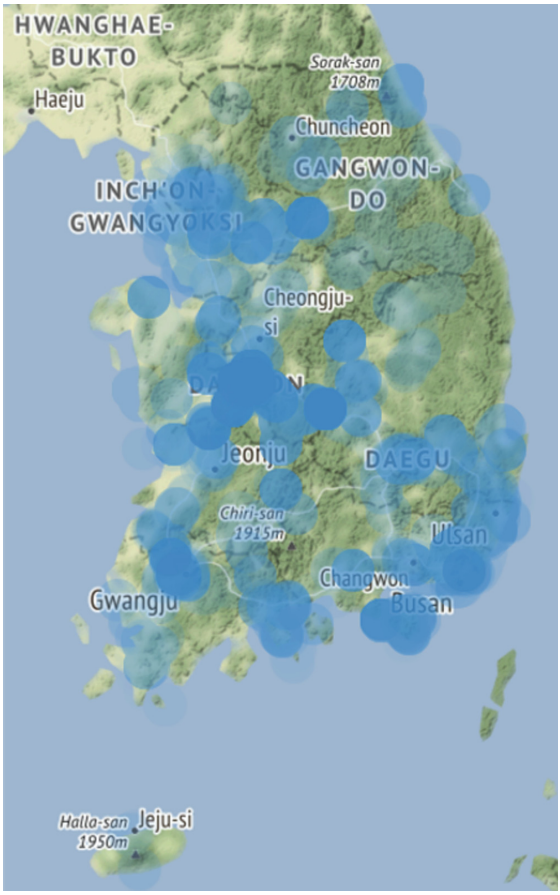


**Fig. 3.** Map with route data

### 3.3 Experimental Results About LCRS Data

Figure 4 shows the result of LCRS algorithm with 10% of raw source data. With LCRS algorithm, highly populated areas are clearly identified. With these result, we can identify highly crowded routes.

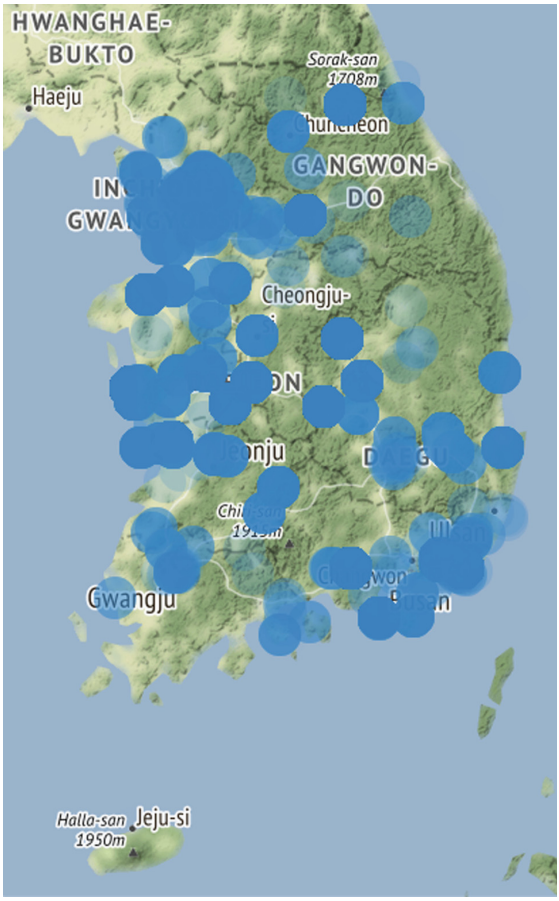


**Fig. 4.** Map of results from LCRS

### 3.4 Voting Results About LCRS Data

Figure 5 show the result of voting from the result of LCRS algorithm. By increasing of radius of candidate area, we identified notable locations more clearly. The most crowded route is Gyeongbu expressway from Busan to Okcheon hub. Three most populated locations are Seoul, Busan and Okcheon hub.

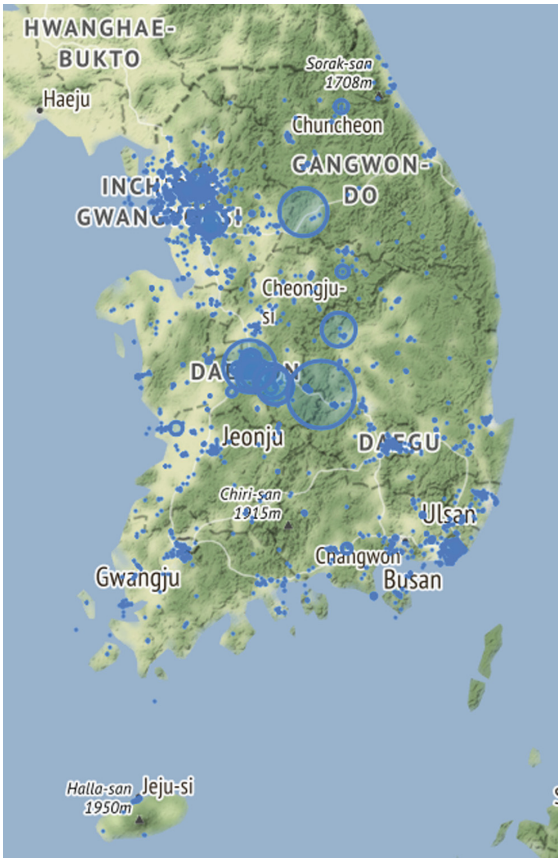


**Fig. 5.** Map after voting

### 3.5 Candidate Location for Logistic Hub

However, from the results shown in Fig. 5, there are another candidate area as shown in Fig. 6 where area A stands for 'Okcheon', B stands for 'Geumsan', C stands for 'Boeun', and D stands for 'Moongyeong'.




**Fig. 6.** Map with candidate locations for logistic hub

## 4 Conclusions and Future Research

$$actual_distance = \sum Distance(start, Hub) + Distance(Hub, end) \quad (1)$$

$$candidate_distance = \sum Distance(start, candidate) + Distance(candidate, end) \quad (2)$$

$$actual_time = \sum Time(start, Hub) + Time(Hub, end) \quad (3)$$

$$candidate_time = \sum Time(start, candidate) + Time(candidate, end) \quad (4)$$

After these analysis by proposed process, the best hub candidates can be determined by calculating economical benefits from the distance and time perspective of each hub candidate [8]. actual_distance (Eq. 1) is expression to calculate actual delivery distance with current Okcheon hub. actual_time (Eq. 2) is expression to calculate actual delivery time with current Okcheon hub. candidate_distance (Eq. 3) is expression to calculate actual delivery distance with new candidate hubs. candidate_time (Eq. 4) is expression to calculate actual delivery time with new candidate hubs. Based on the equations, Algorithm 3 shows the process of benefit calculation for candidate hubs. Candidate hub is one of Boeun, Moongyeong, and Geumsan. Hub stands for current existing hub, Okcheon. Table 5 shows calculated values for economic benefits with every hubs. Distances is presented in kilometers and time is presented in minutes. As a result, once we place hub at Geumsan we can have reduced total distance of 50,000 km and reduced time of 20,600 min total 20,000 parcels delivered.

**Table 5.** Benefits on transportation distance with hub candidates

| Hub candidate | Actual distance | Candidate distance | Distance benefit | Actual time | Candidate time | Time benefit |
|---|---|---|---|---|---|---|
| Okcheon | 137.78 | - | - | 110.19 | - | - |
| Boeun | - | 141.76 | −3.98 | - | 113.44 | −3.25 |
| Moongyeong | - | 146.01 | −8.23 | - | 121.40 | −11.21 |
| Geumsan | - | 135.28 | 2.5 | - | 109.16 | 1.03 |

**Algorithm 3.** Bebefit Calculation

**Input:** A Set of origin source data
**Output:** A Set of distance and time data about candidate hub
**PreProcessing(***conclusions***)** {
1: **for** (i=0; i < len(*InputDataSet*); i++) **do** // len(*A*) = length of array A
2: **if** name == 'Okcheon' **then**
3: name = 'candidate hub address'
4: inputDB.execute("select Distinct(pathnum) from position where name='Okcheon'") // inputDB = input database
5: SelectedPathnum = inputDB.fetchall()
6: USE route crawling program Where pathnum = Selected Pathnum That Used 2nd process
7: **end if**
8: **end for**
9: **return** [Conclusions]
10: }

It is reported that the biggest parcel courier handles 1,000,000,000 parcel delivery for one year. Considering this huge number of delivery, the national

economy can have big benefit with the construction of new hub. Of course, another considerations must be done from the aspect of new construction.

In addition, we can have more detailed result with our method once we can obtain the whole data owned by logistics companies. Even though the parcel delivery system is applied in Republic of Korea, the similar methods can be applied to other countries. Another point of future research is application of parallel computing techniques such as GPGPU, since we spend much of time to finish every process step with 100,000 raw data. It is desirable to reduce computation time in order to repeat our process for new data set or for data sets of other nations.

## References

1. Deltas, G., Desmet, K., Facchini, G.: Hub-and-spoke free trade areas: theory and evidence from Israel. Can. J. Econ. **45**, 942–977 (2012)
2. Andriamananjara, S., Das, G.G.: Hub-and-spokes free-trade-agreements in the presence of technology spillovers: an application to the western hemisphere. USITC Office of Economics Working Paper No. 2004–09-A, 13 September 2004
3. Wonnacott, R.J., Kowalczyk, C.: Hubs and spokes, and free trade in the Americas. National Bureau of Economic Research 1050 Massachusetts Avenue Cambridge, MA 02138, October 1992
4. Pang, J., Ma, Y., Wang, X.: Research on hub location and routing distribution for hub-and-spoke logistics network. In: 6th International Conference on Machinery, Materials, Environment, Biotechnology and Computer, January 2016
5. Chen, H.-M., Chou, Y.-C., Chen, Y.-H.: Pickup and delivery routing with hub transshipment across flexible time periods for improving dual objectives on workload and waiting time. Transp. Res. Part E Logist. Transp. Rev. **61**, 98–114 (2014)
6. Hakonen, H., Bergroth, L., Raita, T.: A survey of longest common subsequence algorithm. In: Seventh International Symposium on String Processing and Information Retrieval, SPIRE 2000. IEEE, 06 August 2002
7. Deorowicz, S.: Quadratic-time algorithm for a string constrained LCS problem. Inf. Process. Lett. **112**(11), 423–426 (2012)
8. Kim, N.-S., Lee, E.-J.: Development of public transport competitiveness indicator using web maps comparison of travel time and distance between public transportation and passenger cars. Korean Soc. Transp. **73**, 536–566 (2015)