한국ITS학회논문지 J. Korea Inst. Intell. Transp. Syst.

전략적 정보제공을 위한 침수영향구역 클러스터링

Dentifying and Clustering the Flood Impacted Areas for Strategic Information Provision

박 은 미* · 빌랄 무하메드**

* 주저자 및 교신저자 : ㈜데이터위즈 대표/목원대학교 도시공학과 교수
** 공저자 : ㈜DP World Business Intelligence team Developer

Eun Mi Park* · Muhammad Bilal**

* CEO, Datawiz Inc, Professor Deepartment of Urban Engineering Mokwon University
** Developer, Business Intelligence team, DP World, Inc

† Corresponding author : Eun Mi Park, Peunmi@mokwon.ac.kr

Vol.20 No.6(2021)
December, 2021
pp.100~109

pISSN 1738-0774
eISSN 2384-1729
https://doi.org/10.12815/kits.2021.20.6.100

Received 10 November 2021
Revised 25 November 2021
Accepted 29 November 2021

요 약

본 연구는 폭우로 인해 도로침수가 발생되고 그로 인한 교통상황 악화가 발생할 때, 도로이용자와 침수와 혼잡 상황을 관리하는 시의 관리자들에 필요한 정보를 생산하기 위한 방법론에 대한 연구이다. 홍수와 같은 재난상황에서, 도로이용자들의 2차 피해를 막고, 도로상황 악화를 방지하며 빠른 회복을 위해서는, 적절한 정보가 제공되어야 한다. 도시의 규모에 따라 차이가 있겠으나, 도시에 수천 개의 구간이 존재하고, 특히 홍수와 같은 상황에서 수백 개 내지 천개 이상의 혼잡구간이 존재할 때, 개별 구간단위 혼잡수준 정보는 재난상황관리에 더 이상 유용하지 않다. 본 연구에서는 홍수상황에 영향을 받는 링크들을 공간적으로 클러스터링하고, 클러스터에 포함되지 못하는 영향 링크들은 정보제공 대상에 열외 시켜 무의미한 정보는 제외될 수 있도록 하였다. 또한 클러스터의 시공간적 특성, 즉 시간적 지속성, 공간적 크기를 산정하여, 영향 지역의 심각도 정보가 제공될 수 있도록 하였다. 본 연구를 통하여 만들어진 정보는 도로 이용자와 도시 관리자 모두가 홍수로 파급된 도로네트워크 문제에 적절히 대응하게 하는데 활용될 수 있을 것으로 기대된다.

핵심어 : 혼잡링크, DBSCAN, 도로 문제 구역, 홍수 영향 구역, 링크 클러스터링

ABSTRACT

Flooding usually brings in disruptions and aggravated congestions to the roadway network. Hence, right information should be provided to road users to avoid the flood-impacted areas and for city officials to recover the network. However, the information about individual link congestion may not be conveyed to roadway users and city officials because too many links are congested at the same time. Therefore, more significant information may be desired, especially in a disastrous situation. This information may include 1) which places to avoid during flooding 2) which places are feasible to drive avoiding flooding. Hence, this paper aims to develop a framework to identify the flood-impacted areas in a roadway network and their criticality. Various impacted clusters and their spatiotemporal properties were identified with field data. From this data, roadway users can reroute their trips, and city officials can take the right actions to recover the affected areas. The information resulting from the developed framework would be significant enough for roadway users and city officials to cope with flooding.

Key words : Congested links, DBSCAN, Disruptive network, Flood Impacted Areas, Link clustering

Ⅰ. Introduction

Disruption of transportation networks has been widely covered in the past decade. Flooding is one of such event which results in the disruption of transportation networks. This disruption is in form of congestion and the blocking of road links due to partial flooding. In transportation network, it appears in the form of 'snakes'-a series or group of congested links- which increase of decrease in size over time interval. These snakes either split up or merge with neighbor snakes. However, the area where snake is formed and dispersed is usually static as snakes formation and deformation often resides in recurrent congestion of normal day. If someone goes into the area surrounded by several snakes, they are stuck in the traffic. Additionally, the severity of disruption induced by different areas covered by snakes are not similar in nature. Certain parts of transportation network may be more important due to amount of traffic they carries, the centralities they got, the spatio-temporal coverage they have, and etc.. Thus, there is a need to identify these impacted area and their severity. Identification of such impacted clusters can help maneuvering flooding-reactive actions such as emergency evacuation, detouring information provision for roadway users, and etc..

This study employed a simplified methodological framework to determine those affected areas and rank them using spatial and temporal factors and congestion severity of the areas. This study used DBSCAN algorithm to identify the neighboring impacted links as a cluster and used Convex Hull to make boundary of the cluster. Over the last few decades, various approaches have been used in transportation researches to identify critical links. However, the past studies were too cornered by individual links in the networks. They ignored the scenario that if an link is congested, the neighbor links are affected. And therefore, traveler should also informed to avoid the links surrounded by congested links. Identification of snakes, rather than individual link congestion, followed by clustering the neighboring snakes aims to fill this gap in the past studies.

As an attempt to explore the likely impact of flood on the city, this study aims to: (1) Identify the critical links (2) cluster the neighboring links (3) rank the clusters by spatio-temporal congestion severity factor. This study used real road network data of Daejeon. Algorithms used for clustering and ranking are DESCAN, gift wrapping and K-means clustering, respectively. Congestion analysis is performed based on the speed data obtained from OBU(On-Board Unit) and RSE(Roadside Equipment) Communication in Daejeon roadway network.

This paper is organized as follows: In Section 2, briefly go through the literature review results. Then data description and methodological framework are followed in Section 3. Then research results are presented in Section 4 and 5. And finally conclusions are drawn in Section 6.

Ⅱ. Literature Review

Measures of link ranking have been studies for a long period of time with an emphasis on criticality of the network disruption and vulnerability. Most of the studies found in the literature measure the importance of the link due to disruptions in order to measure network vulnerability (Jenelius and Mattsson, 2012; Rupi et al., 2014). Similarly, there are various measures proposed in the past that calculate the link criticality in a network (Rodriguez-Nunez and Garcia-Palomares, 2014; Sullivan et al., 2010). Some of the measures adopted in past studies

have been found to compute the accessibility of a link in a network (Luathep et al., 2013). Some studies use link efficiency in a network and others evaluate link interuptions and alternate routing (Qiang, 2007). Some studies have also attempted to rank the links either based on combination of a different criteria by using spatiotemporal patterns of alternative travel paths. Morever, researchers have also attempted to conduct studies related to evaluation of infrastructure looking at the broader impacts including infrastructure disruptions and accessiblity index (Hasan and Foliente, 2015).

Most of the previous studies focused on individual links. Studies related to identifying the impacted areas and the ranking has been missing so far. This study is undertaken to fill the gap.

Ⅲ. Methodology

1. Data site

Travel speed data of Daejeon used in this study were collected from RSE (Roadside Equipment)-OBU(On-board Unit) communications. The link speed data are obtained from Dajeon ITS center DB. There is also the predefined road network data available, which includes geographic information of intersections and streets. 15min link speed data on both flood and normal day were used for this study. Basic information for each link has defined such as starting node name, ending node name, direction of the route on which certain link lies, and etc.. Data preprocessing was carried out in two steps. In the first step, we made a sample out of the whole day data. Knowing that congestion varies across time of day as well as by the direction of traffic, we sampled the data in two groups i.e., morning peak hours 7-9 am and afternoon peak hours 6-8 pm. In the second step we defined the parameter to distinguish whether congestion exists or not. In order to characterize the link state, the parameters related to the traffic state need to be analyzed and processed accordingly. Commonly used parameters to characterize the traffic state are flow and speed. This paper select speed data to analyze the traffic condition of the road network.

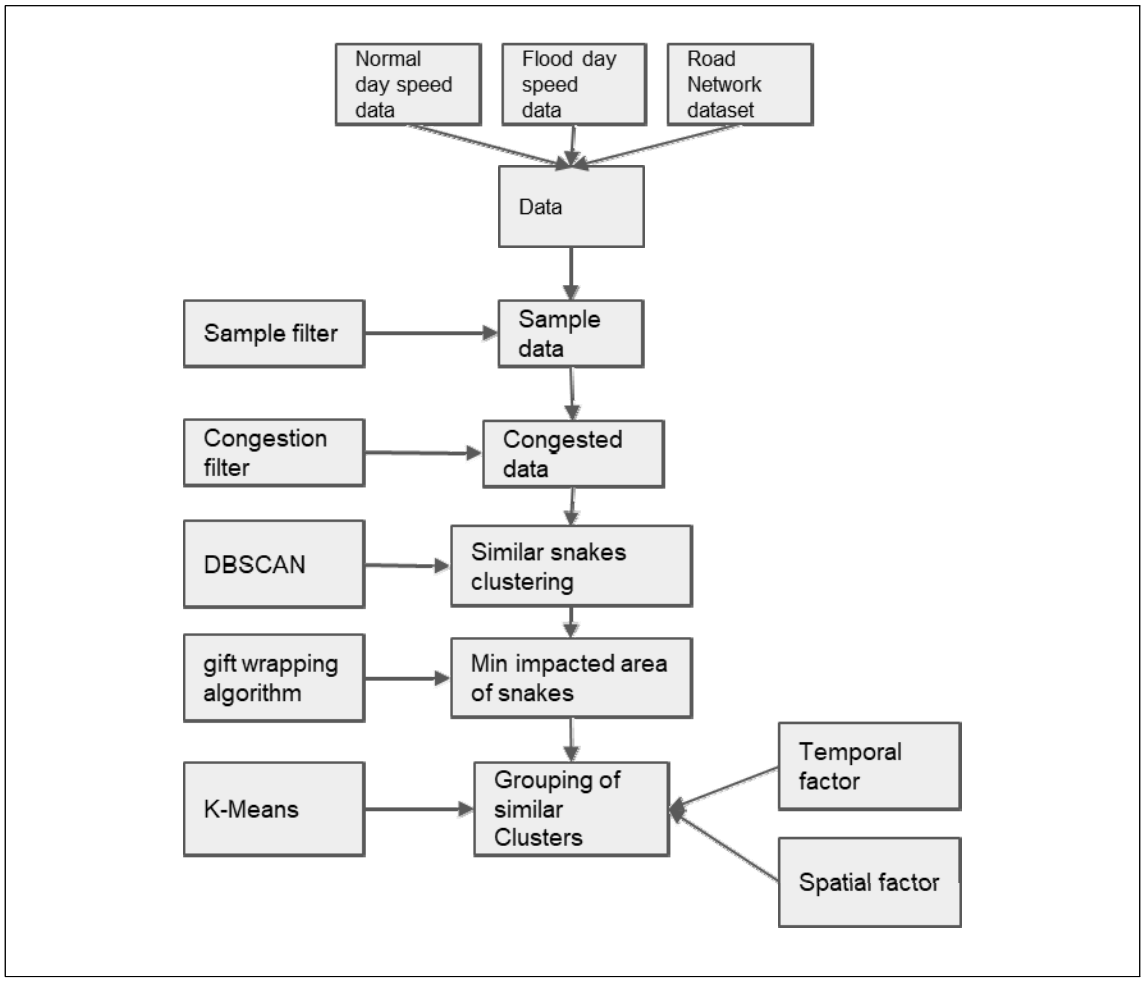
In order to define the criteria for certain link to be impacted by flood, we used the speed difference of a link between the normal and the flood day, making sure that the same link was not congested on the normal day. Let $\overline{V_n}$ speed on normal day and $\overline{V_f}$ speed on flood day, then a link is considered impacted if following condition hold.

$$\overline{V_f} < 10KM/hr \text{ and } \overline{V_n} > 10KM/hr \text{ and } \overline{V_f} - \overline{V_n} < -5 \quad \cdots\cdots (1)$$

2. Framework Development

In this section we will describe the methodological framework developed (refer to <Fig. 1>). We took sample of the whole data set in order to analyze the data. As mentioned before, the sampling was based on time interval, direction of traffic and speed level. Then, we used DBSCAN algorithm to identify the neighbor congested snakes and clustered them together. Next step was to find the boundary of the impacted area by making a convex hull

around the obtained clusters using gift wrapping algorithm. After finally getting the clusters, we grouped them using spatial and temporal factors. We used number of congested links within the clusters as spatial factor and duration for which links within cluster stayed as temporal factor. To accomplish grouping similar valued clusters, K-means algorithm was used.



<Fig. 1> Methodological Framework

Ⅳ. Clustering

Clustering represents the most commonly used and more powerful unsupervised learning mechanism in machine learning. It is a useful tool that aims to classify the input data into sets depending on some similarity calculations. There algorithms are categorized into groups like partition algorithms, density-based algorithms, hierarchical algorithms (Ester et al., 1996). Among them DBSCAN has been selected for our proposed research framework

because it has many features that make it suitable compared to other clustering algorithms. DBSCAN is an effective density-based clustering algorithm for spatial data systems due to its ability to discover clusters with arbitrary shapes. The grouping process with DBSCAN can be described as a tree. It starts with any point that has at least the MinPts-a given parameter- closest points within a given radius ε. Then do the first search along each of these closest points by checking how many points there are within the radius ε. If it has at least MinPts neighbors, then that point becomes a branch, and all its neighbors are added to the cluster. As our goal is to identify clusters while removing the outliers i.e., congested links not close enough to other links as well as the other clusters.

DBSCAN is performed based on two attributes (latitude and longitude) of the congested links initially obtained using simple filtering process. It requires two parameters. One is the epsilon ε, which specifies how close points should be to each other to be considered a part of cluster. The other is MinPts, which specifies how many neighbors a point should have to form a cluster. Selection of the ε and MinPts is key to the DBSCAN algorithm (Rahmah and Sitanggang, 2016).

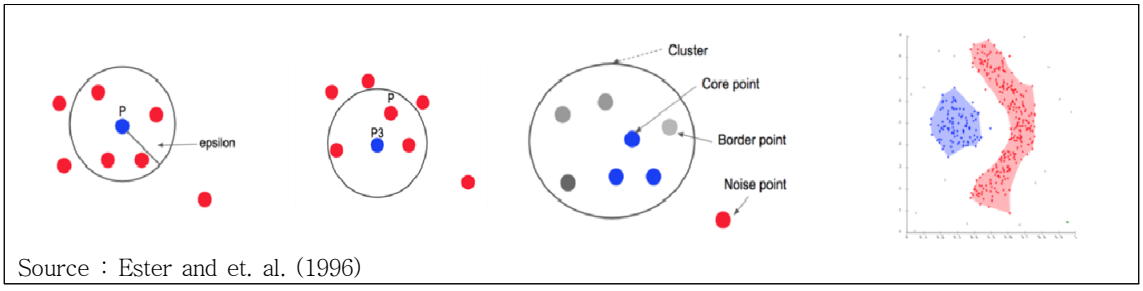
For a given neighborhood distance ε and the minimum number of neighbors, MinPts, the algorithm flow is as follows. When N coordinates of each point are represented as D = (x1,x2,x3,.....,xn), the Euclidean distance between two difference points is calculated as

$$dist(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad \cdots\cdots (2)$$

For fixed ith point, points that satisfy the following condition are grouped into one cluster, which is expressed as

$$dist(X,Y) \leq \varepsilon \ (\text{for } i \neq i') \quad \cdots\cdots (3)$$

where ε is the set threshold radius.



Source : Ester and et. al. (1996)

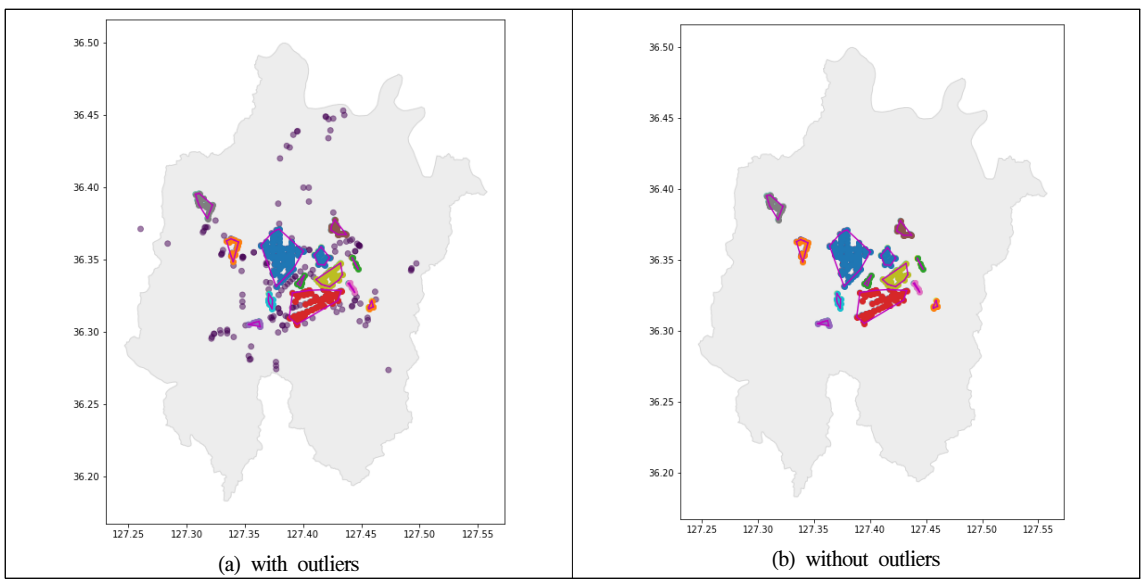
<Fig. 2> Concept of DBSCAN – Core, Border and Noise points

To find a suitable value for epsilon, the distance to the nearest n points for each point is calculated and sorted. And the results are plotted. Then the point where the slope change is most pronounced, is selected as epsilon value. The distance from each point to its closest neighbour was calculated using the NearestNeighbors (Fukunaga and Narendra, 1975). The k neighbors method returns two arrays, one which contains the distance to the closest

neighbor points and the other which contains the index for each of those points. In the next step, the results were sorted and plotted. The optimal value for epsilon would be found at the point of maximum curvature. It was assumed that the area occupied would be significant if there are more than 6 links close to each other. If the links obtained are not close enough to each other and do not form a snake larger than 6 links combined together, they are considered as outlier and/or insignificant areas. For the clusters obtained, a convex hull was created around the links within a cluster to find the minimum area impacted by the flood. After the impacted area identified, the congested time repetition of each link and the number of links within each cluster were calculated.

Ⅴ. Results

Results of morning peak hours 7-9am after flooding, are shown in Figure 3 with outliers and without ouliers. Using k-nearest neighbor analysis with minimum 6 points as input parameter, the optimal value of epsilon came out to be 0.5 km. So, in the DBSCAN algorithm, epsilon and MinPts value 0.5 km and 6 are used, respectively.



(a) with outliers (b) without outliers

<Fig. 3> Impacted areas, 7-9am

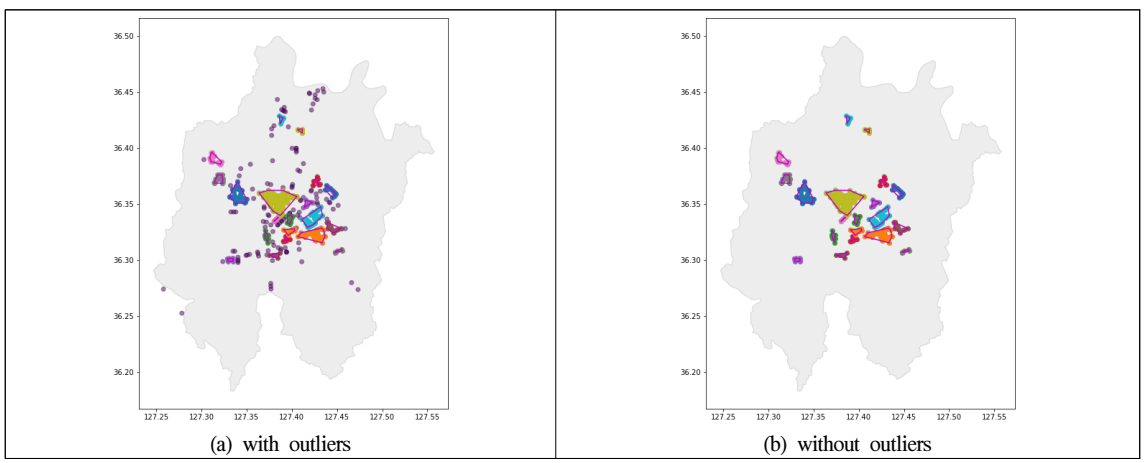
13 clusters are obtained. Cluster along with its spatial and temporal properties are shown as <Table 1>. In this table, the outliers denoted by －1 contain 144 links. The author wants to point out that the outliers do not necessarily mean that they are not really congested or impacted. Rather the outliers are not significant enough for citizens to be informed about. This research considers only impacted areas, i.e., clusters of neighboring links significant, whose information, in turn, should be conveyed to citizens. In this context, links spreaded out and/or isolated were considered outliers and filtered out.

<Table 1> Clusters and groups, 7-9am

Clusters	Average Repetition	Number of links
-1	4.03	144
0	3.44	140
1	3.35	20
2	4.78	9
3	3.13	70
4	4.25	8
6	3.00	6
5	2.93	15
7	3.54	24
8	4.55	31
9	4.86	7
10	3.94	16
11	3.00	6
12	4.83	6
Groups	**Average Repetition**	**Average Number of links**
1	3.43	140
2	3.19	43.78
3	4.46	18.79

After identifying the clusters and impacted areas, this research grouped the clusters with similar temporal and spatial properties using k-mean clustering. 3 groups are obtained and shown in <Table 1>. Group 1 was the biggest cluster with 140 link.

Similarly, the results for the afternoon peak 6-8pm after flooding were obtained, which are shown in <Fig. 4> and <Table 2>. Using k-nearest neighbour analysis with minimum 6 links as input parameter, the optimal value of epsilon came out to be 0.5 km, which is the same as the one of morning peak period.



(a) with outliers (b) without outliers

<Fig. 4> Impacted areas, 6-8pm

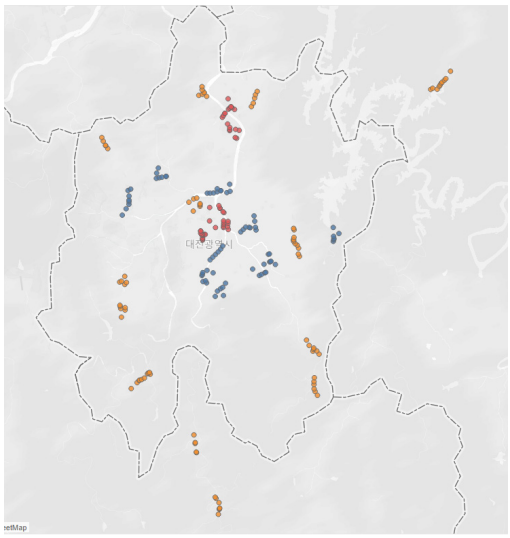
<Table 2> Clusters and groups, 6-8pm

Clusters	Average Repetition	Number of links
-1	3.69	124
2	3.33	9
6	4.50	6
8	4.19	108
0	3.78	49
1	3.11	36
3	3.00	6
4	5.71	7
13	4.33	9
5	4.43	14
7	4.20	15
15	1.29	7
9	3.13	31
10	4.00	16
11	3.25	8
12	3.57	7
14	5.50	6
16	4.12	16
17	3.50	6
18	4.00	6
19	5.86	7
Groups	**Average Repetition**	**Average Number of links**
1	3.07	23.38
2	4.26	23.98
3	4.20	108

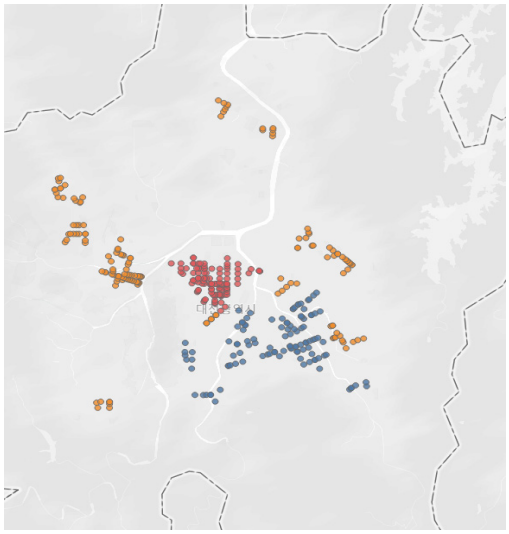
The number of clusters obtained were 19. Clusters and their spatial and temporal properties are shown in <Table 2>. Adopting similar methodological framework, the clusters were grouped with similar temporal and spatial properties using K-means clustering and 3 groups obtained as shown in <Table 2>. Group 3 was the biggest of clusters with 108 links. The groups for morning and afternoon peaks are shown in <Fig. 5> and <Fig. 6>, respectively.

Ⅵ. Conclusions

Flooding usually brings on disruptions and aggravated congestion to the roadway network. Right information should be provided for road users to avoid the flood impacted areas and for city officials to endeavor to recover the network. However, the information about individual link congestion may not be conveyed on roadway users and also not helpful to city officials in that there are too many links congested.



<Fig. 5> Groups, 7-9am



<Fig. 6> Groups, 6-8pm

More significant information may be desired especially for a disastrous situation, such as 1) which places to avoid during flooding 2) which places are feasible to drive on before flooding occurs. Furthermore, such additional information as how much risk the road users might take if they stick to their route and pass through the impacted areas, might also be desired. This paper aims to develop a framework to identify the flood impacted areas and their criticality. It is expected that the research results should be significant enough for roadway users to avoid aggrevated congestion and for city officials to cope with flooding.

This study has certain limitations as well. The filtering criteria used to identify the impacted links can further be improved by inclusion of other factors such as traffic volume and capacity of the link. Selecting adequate clustering thresholds, ε and MinPts, is a crucial issue. This issue cannot be addressed only by theoretical and/or technical experiments but by extensive field experiments, which is undergoing as a next stage research. Ranking the obtained clusters in terms of criticality by adding additional spatial and temporal factors also remains further study.

ACKNOWLEDGEMENTS

본 연구는 한국산업기술평가관리원 지원의 『지역맞춤형 재난안전 문제해결 기술개발 지원 사업』의 연구 성과입니다.

REFERENCES

Ester M., Kriegel H. P., Sander J. and Xu X. et al.(1996), “A density-based algorithm for discovering clusters in large spatial databases with noise,” *Kdd*, vol. 96, pp.226-231.

Fukunaga K. and Narendra P. M.(1975), "A branch and bound algorithm for computing k-nearest neighbors," *IEEE Transactions on Computers*, vol. 100, no. 7, pp.750-753.

Hasan S. and Foliente G.(2015), "Modeling infrastructure system interdependencies and socioeconomic impacts of failure in extreme events: Emerging R&D challenges," *Natural Hazards*, vol. 78, pp.2143-2168.

Jenelius E. and Mattsson L. G.(2012), "Road network vulnerability analysis of area-covering disruptions: A grid-based approach with case study," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 5, pp.746-760.

Luathep P., Suwansunthon A., Suttiphan S. and Taneerananon P.(2013), "Flood Evacuation Behavior Analysis in Urban Areas," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 13, pp.178-195.

Qiang A. N.(2007), "A network efficiency measure for congested networks," *Europhysics Letters Association*, vol. 79, no. 3, 38005.

Rahmah N. and Sitanggang I. S.(2016), "Determination of optimal epsilon value on DBSCAN algorithm to clustering data on Peatland hotspots in sumatra," *In IOP Conference Series: Earth and Environmental Science*, IOP Publishing, vol. 31, 012012.

Rodríguez-Núñez E. and García-Palomares J. C.(2014), "Measuring the vulnerability of public transport networks," *Journal of Transport Geography*, vol. 35, pp.50-63.

Rupi F., Bernardi S., Rossi G. and Danesi A.(2014), "The Evaluation of Road Network Vulnerability in Mountainous Areas: A Case Study," *Networks and Spatial Economics*, vol. 15, pp.397-411.

Sullivan J. L., Novak D. C., Aultman-hall L. and Scott D. M.(2010), "Identifying critical road segments and measuring system-wide robustness in transportation network with isolating links: A link-based capacity-reduction approach," *Transportation Research Part A*, vol. 44, pp.323-336.