



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

박 종 수 교수지도
석사학위 청구논문

수도권 교통 흐름의
지역적 연계 구조 분석

2011

성신여자대학교 교육대학원
교육학과 전자계산교육전공
소 주 영

수도권 교통 흐름의 지역적 연계 구조 분석

박 종 수 교수지도

이 논문을 석사학위논문으로 제출함

2011년 5월

성신여자대학교 교육대학원
교육학과 전자계산교육전공
소 주 영

인 준 서

소주영의 석사학위 논문으로 인준함.

심사위원 _____ (인)

심사위원 _____ (인)

심사위원 _____ (인)

성신여자대학교 교육대학원

논문 개요

서울 수도권 지역에서 교통 혼잡의 문제는 지속가능한 도시를 위해서는 아주 중요한 문제다. 따라서 본 논문에서는 수도권 지하철을 이용하는 승객의 승, 하차의 흐름을 분석하기 위하여 각각의 역이 아닌 가까운 거리에 위치한 역을 그룹화 하여 지역적으로 승객의 흐름을 알아보았고, 분류 방법에 따른 그룹화의 결과 비교를 위해 분류 방법을 세 가지 방법을 통해 분석해 보았다.

수도권 지역에서 지하철 승객 흐름을 분석하기 위하여 지하철역들을 몇 개의 그룹으로 군집화 하기 위해 고려한 세 가지 방법으로 첫째, 한강을 기준으로 분석하였고 둘째, K-means 알고리즘을 이용하여 지하철역들을 군집화 하였다. 마지막으로 MST(Maximum flow Spanning Tree) 방법을 적용하여 시간대별 승객 흐름을 분석하였다. 또한 세 가지 방법에 따른 결과를 시각화 하여 분석하기 위해 군집화된 지하철역들 사이의 승객 흐름과 연계 구조는 서로 다른 색깔로 표현하여 승객 흐름의 양을 보여주는 간선들과 노드들로 구성된 방향 그래프로 나타내었다.

본 논문 결과의 분석을 통하여 같은 그룹 내에서의 이동이 매우 활발하며, 아침시간대와 저녁시간대에 붐비는 그룹은 대부분이 회사가 밀집되어 있는 지역임을 알 수 있었다. 이 논문의 결과는 대중교통의 서비스 개선에 기여할 수 있고, 지하철 역 그룹들 사이의 관련성 및 연계적인 구조는 해당되는 지역 개발을 위한 토지 이용에 적용될 수 있을 것이다.

목 차

논문개요

I. 서론	1
II. 관련연구	3
III. 교통카드 트랜잭션을 이용한 수도권 지하철 분류 방법	7
1. 교통카드 데이터 구성	7
2. 분류 방법	9
2.1. 한강을 기준으로 그룹으로 분류	10
2.1.1 분류 단계	10
2.1.2 Floyd Shortest Path 알고리즘	11
2.1.3 Floyd Shortest Path 알고리즘을 통한 데이터 분석	12
2.2. K-means 알고리즘을 이용하여 그룹으로 분류	14
2.2.1 분류 단계	14
2.2.2 K-means 알고리즘	15
2.2.3 K-means 알고리즘을 이용한 데이터 분석	16
2.3. MST 기준 시간대별 그룹으로 분류	17
2.3.1 MST(Maximum flow Spanning Tree)를 이용한 분석	18

2.3.2 분류 단계	18
2.3.3 MST(Maximum flow Spanning Tree)를 이용한 지하철 역 분류 방법	19
IV. 실험결과 및 분석	21
1. 실험환경	21
2. 실험결과 및 분석	22
2.1 한강을 기준으로 그룹으로 분류	22
2.2 K-means 알고리즘을 이용한 그룹 분류	26
2.3 MST 기준 시간대별 그룹 분류	32
2.4 각 분류 방법의 비교 분석	49
V. 결론	50

참 고 문 헌

ABSTRACT

표 목 차

[표 1] 교통카드 트랜잭션의 속성과 예제	9
[표 2] 하루 동안의 시간대별 승객 흐름으로 분류한 역	19
[표 3] 1, 2단계의 각 그룹 역 개수	22
[표 4] 1, 2단계 지하철 승객 수	26
[표 5] K-means 알고리즘으로 분류한 단계별 그룹 개수	27
[표 6] K-means 알고리즘으로 분류한 단계별 승객 수	31
[표 7] MST 기준 시간대별 8개의 그룹 역 개수	33
[표 8] MST 기준 시간대별 4개의 그룹 역 개수	34
[표 9] MST기준 시간대별 분류를 통한 하루 동안의 각 그룹 역 개수 ..	42
[표 10] MST기준 시간대별 분류 및 전체 8개 그룹의 승객 수	47
[표 11] MST기준 시간대별 분류 및 전체 4개 그룹의 승객 수	48
[표 12] 분류 방법 비교 분석표	49

그림 목 차

[그림 1] 오전 9시 승하차 인원의 시각화	6
[그림 2] Floyd Shortest Path Algorithm	11
[그림 3] 4개의 그룹으로 분류하기 위한 알고리즘의 예제	13
[그림 4] K-means 알고리즘의 예	15
[그림 5] K-means 알고리즘을 이용하여 역 분류를 위한 순서도	16
[그림 6] 서울 수도권 지하철 시스템에서 승객 흐름에 대한 Maximum flow Spanning Tree(MST)	18
[그림 7] 지하철 노선도 상에서 역 분포	23
[그림 8] 1단계 그룹 승객 흐름도	24
[그림 9] 2단계 그룹 승객 흐름도	25
[그림 10] K-means 알고리즘을 이용하여 분류한 역 분포	28
[그림 11] random 중심점의 승객 흐름도	29
[그림 12] 노선도 양 끝에 위치한 중심점의 승객흐름도	30
[그림 13] MST 기준 오전시간대 그룹 역 분포	35
[그림 14] MST 기준 낮 시간대 그룹 역 분포	36
[그림 15] MST 기준 저녁 시간대 그룹 역 분포	37
[그림 16] MST 기준 아침시간대 승객흐름도	37
[그림 17] MST 기준 낮 시간대 승객흐름도	39
[그림 18] MST 기준 저녁시간대 승객흐름도	40
[그림 19] MST 기준 하루 동안의 기준 역을 분류하기 위한 순서	41
[그림 20] MST 기준 하루 동안의 그룹 역 분포	43
[그림 21] MST 기준 하루 동안의 승객 흐름도	44
[그림 22] MST 기준 시간대별 승객 흐름도	46

I. 서론

최근 우리나라의 수도권 내 교통 문제는 날로 심각해져가고 있다. 따라서 대중교통의 중요성이 부각되어 교통정책에서는 자가용을 억제하여 교통 혼잡을 줄이고 대중교통의 활성화를 위해 대표적으로 수도권 광역교통체계와 간선급행버스(BRT: Bus Rapid Transit)사업을 추진하였다[1].

그러나 이러한 노력에도 불구하고 지하철의 분담률은 오히려 줄어들었으며, 자가용 이용률은 큰 차이가 없는 것으로 나타났다[3]. 지하철 이용률은 10년 전보다는 높아졌지만 대중교통체계가 개편된 후인 2004년부터는 미세하게 하락하는 추세로써 지하철의 이용률을 크게 높이는 것에는 실패하였다[2]. 따라서 현재 수도권에서 대중교통의 이용률을 높이기 위해서 대중교통을 이용한 연구가 시급하다. 이것에 관한 연구[3]은 수도권의 대중교통체계의 개편에 따른 대중교통 이용객 현황을 분석하였다. 그 결과 전체 대중교통의 이용객은 증가하고 있으나 버스와 지하철을 비교해 볼 때 지하철 승객수는 감소하고 있으므로 이에 따른 대책 마련이 필요하다고 설명했다. 또 다른 연구[4]는 지하철을 이용하기 위한 연계수단으로 가장 큰 비중을 차지하고 있는 수단은 도보 이용(79.06%), 다음으로 일반버스(9.50%), 마을버스(6.91%)순으로 나타났다. 지하철의 승객수를 늘리기 위해서는 도보 이용에 더불어 버스와 지하철의 효과적인 연계 방식의 연구가 필요하고, 결과적으로 대중교통의 높은 이용으로 교통 혼잡을 줄일 수 있다.

대중교통을 연구하기 위하여 가장 중요한 자료인 교통카드는 현 수도권 내에서 스마트카드를 사용하여 대중교통을 이용하는 통행자의 기록이 담겨 있다. 각 트랜잭션을 분석하면 탑승유형, 통행패턴, 등의 정보를 얻어낼 수 있다. 따라서 최근 교통카드의 데이터베이스를 기반으로 데이터마이닝의 기법을 적용하고 지리학적으로 분석하여 수도권 지하철의 통행 패턴이나 통행

유형에 대한 선행연구가 이루어지고 있다[5-10].

이에 본 논문에서는 교통카드의 데이터자료를 이용하여 수도권 지하철의 승객 흐름을 분석하고자 수도권 지하철의 주요 역을 선별하여 이에 따라 모든 역을 그룹화 하여 연구하였다. 모든 역을 그룹으로 나누는 것은 특정 역이 아닌 그 주위의 비슷한 위치에 있는 역을 묶어 그 지역에 대한 특성을 알아볼 수 있고, 승, 하차가 활발한 그룹 역을 분석하여 지역 간의 연관성을 파악할 수 있으며, 그에 따른 지리학적 특성과 토지 이용 관계를 분석할 수 있다. 이러한 연구를 위해 본 논문에서 가장 중요한 그룹으로 나누기 위한 주요 역 선택의 방법을 세 가지로 나누었다. 첫 번째, 지리학적으로 한강을 기준으로 나누어 임의의 역을 선택하는 방법, 두 번째, K-means 알고리즘을 이용하여 중심 역을 선택하는 방법, 세 번째는 MST(Maximum flow Spanning Tree)기준으로 시간대별로 승객흐름이 많은 역을 선택하는 방법으로 나누어 세 가지 방법으로 인하여 분류된 역의 결과를 비교해 볼 수 있다.

본 논문의 구성은 제 2장에서 본 논문과 관련된 기존의 연구 결과들을 알아보고, 3장에서 교통카드 트랜잭션을 이용한 수도권 지하철의 분류 방법과 사용된 알고리즘에 대해서 설명한다. 4장에서는 각 분류 방법에 따른 실험으로 2005년 6월의 어느 하루 동안의 교통카드 데이터를 적용하여 얻은 결과를 보여주고, 이 결과를 분석하여 시각적으로 비교하여 설명하고, 5장에서 결론 및 향후과제를 설명하며 끝맺는다.

II. 관련 연구

교통카드를 이용하여 승객 흐름을 분석하기 위한 선행연구는 점점 확산되어 가고 있다. 교통카드의 대용량 데이터를 이용하여 승객 흐름을 분석하는 연구에도 어느 곳에 중점을 두느냐에 따라 다양한 연구 결과를 얻을 수 있다. 본 논문에서는 다양한 분석 방법을 크게 세 분류로 나누어 분석해보았다.

첫째, 교통카드 트랜잭션을 통하여 승객흐름을 분석한 연구이다.

대중교통체계가 개편된 이후 수도권 내 대중교통 이용자들은 교통카드를 사용한다. 이 교통카드는 스마트카드를 기반으로 하여 개인당 통행의 기록이 모두 담겨있는 데이터이다. 따라서 이 데이터를 가지고 승객의 흐름을 분석할 수 있다. 교통 카드 내에 담겨있는 데이터에는 다양한 정보(카드 ID, 카드 트랜잭션 번호, 환승횟수, 승차일시, 승차정류장, 하차정류장, 하차일시, 버스노선, 승차금액 등)를 포함하고 있는데, 이러한 데이터는 하루 수천만을 넘고 있어 대용량 데이터에 데이터마이닝 기법을 적용하면 대중교통을 이용하는 승객의 흐름을 분석할 수 있다[6]. [6]의 연구에서는 데이터마이닝 기법을 도입하여 하루 동안의 교통카드 트랜잭션을 분석하였다. 대중교통의 사용자의 통행패턴을 파악하기 위하여 순회패턴을 탐사하는 알고리즘을 적용하여 통행패턴을 찾아내었고, 알고리즘의 성능을 분석하였으며, 공간적인 특성을 분석하였다.

[5]의 연구에서는 [6]의 연구를 확장하여 순회패턴탐사에서 통행 시퀀스를 찾아내고, 이를 바탕으로 통행패턴을 찾아내는 데이터 마이닝 방법의 개발에 초점을 두었다. 그리고 이것을 2004년부터 2006년까지 3년의 각각 다른 시점으로 하루 동안의 교통카드 자료에 데이터 마이닝 기법을 적용하여 정류장 개수, 환승 횟수 등을 비교하고, 통행패턴의 공간적 특징과 시점 간 차

이를 분석하였다.

[1]의 연구에서는 수도권 내 지하철의 승객 통행 패턴뿐만 아니라, 지하철 이용자의 통행연쇄패턴까지 다루어 논리적으로 구분하고, 수도권 광역철도 이용자의 역 별 총 이용 빈도와 시간대별 이용 빈도를 확인하여 구분하고 이에 따른 집단 간의 특성에 대해 분석하였다.

[1],[5],[6]의 연구에서는 승객의 흐름을 교통카드의 대용량 데이터베이스에서 출발역과 도착역을 분석하여 승객의 흐름을 분석하였다면, [10]의 연구에서는 수도권 지하철 교통망에서 승객 흐름을 분석하였다. 각 링크의 승객 흐름을 분석하기 위해서는 수도권 지하철의 교통망을 구성하는 그래프에 대한 자료구조의 분석과 승객의 흐름을 계산하기 위한 알고리즘이 필요하다. 따라서 역과 역 사이에 연결된 링크를 이동하는 승객의 흐름을 분석하기 위해 새로운 방법론을 제시하였고, 교통카드의 대용량 데이터베이스를 대입하여 각 링크마다 시간대별로 승객흐름을 분석하였다.

이 외 다차원 핵심정보배열추출기법을 이용하여 서울시내의 대중교통을 이용하는 패턴과 일반적인 특성을 탐구한 연구도 있다[8].

둘째, 수도권 지하철망의 구조와 토지 이용을 분석한 연구이다.

도시 내에서의 교통흐름은 공간적인 특성과 토지 이용과 밀접한 관계가 있다. 교통흐름은 통행자의 이용형태를 파악할 수 있기 때문에 통행의 목적 등이 나타날 수 있다. 따라서 교통흐름을 분석하여 도시 내의 공간적인 특징을 파악하는 것은 교통뿐 만아니라 지리학적으로 중요한 연구이기도 하다 [11]. 특히 수도권 내 높은 수송률을 차지하고 있는 지하철의 통행 흐름은 통행을 하는 이용자의 목적과 공간적 특성을 밝혀내고, 통행 흐름과 토지 이용 관계를 밝혀낼 수 있다[9]. [9]의 연구에서는 수도권 지하철망 교통흐름의 위상학적 구조를 분석하고 통행 흐름과 토지를 이용하는 관계성을 분석하였다. 특히 지하철 망을 분석하기 위하여 4년 동안 축적된 통행 흐름을

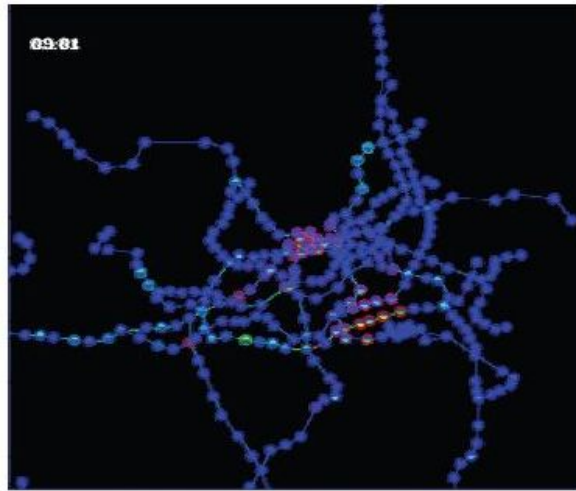
분석하여 위상학적 구조의 변화를 비교하여 분석하였으며, 하루 중 시간대 별 통행 흐름을 알아보기 위해 아침, 낮, 저녁 시간대로 나누어 통행 패턴을 분석하였다. 그리고 지하철 통행 흐름과 지역 간의 관계를 분석하기 위해 상관관계분석과 다중회귀분석을 이용하여 결과를 도출했다.

[7]의 연구에서는 수도권 지하철의 구조적 특성을 연구하기 위해서 지하철 망의 역들을 정점으로 연결되는 link로 구성하여 네트워크 구조로 표현하였다. 지하철망의 네트워크 구조 분석과 역과 역 사이를 실제 거리를 바탕으로 하여 계산되어지는 최단경로 알고리즘을 개발하였다. 이에 대해 실제 통행 기록이 들어있는 교통카드 데이터베이스를 이용하여 통행흐름과 패턴을 비교 분석하였다.

마지막으로, 승객 흐름을 시각적으로 분석한 연구이다.

한 도시 내의 교통흐름을 파악하여 분석하는 것은 결코 쉽지 않다. 이것을 시각화 하여 교통흐름을 본다면 쉽게 이해될 것이다. 교통의 흐름을 시각화 하여 보여준다는 것은 쉽지 않아 국내에선 시각화에 대한 연구가 활발하지 않다.

이에 대한 연구[12]는 수도권 지하철의 교통카드 데이터베이스를 이용하여 통행흐름의 구조를 분석하여 토지이용패턴과 관계성을 시각화하였다. 승객의 흐름을 시각적으로 표현하기 위하여 정보시각화 방법으로 분석하여 역동적으로 보여주었다. 시간의 흐름에 따라 승하차 인원을 시각화하여 나타내었는데 그 중 아래 [그림 1]은 오전 9시의 각 지하철 노선별 승객의 승, 하차 및 역 주변에 머무르는 유동인구 밀도를 시각화 하여 나타낸 것이다 [12].



[그림 1] 오전 9시 승하차 인원의 시각화

본 논문은 이러한 선행 연구를 이어 수도권 지하철의 교통카드 데이터베이스를 이용하여 승객 흐름을 분석하기로 하였다. 각각 다른 방법으로 지하철의 주요 역을 구하고 그 중심 역으로 인한 모든 역들을 그룹으로 나눈 후, 서울시내의 지역구분으로 승객의 흐름을 알아보고, 시간대별 유동인구가 가장 많은 지역의 분석과 시각적인 그림을 통하여 검증할 것이다.

Ⅲ. 교통카드 트랜잭션을 이용한 수도권 지하철 분류 방법

교통카드 트랜잭션에는 대중교통을 이용하는 승객의 교통 정보가 포함되어 있어 교통 카드를 분석하면 지하철을 이용하는 승객들의 흐름을 알 수 있다. 본 논문에서는 수도권 지하철의 역들을 그룹으로 나누어 승, 하차가 활발히 이루어지는 역과 지역적인 특성을 분석하여 관계성을 파악한다. 또한 시간대별로 이루어지는 승객의 승, 하차의 분석을 통하여 시간대와 지역적인 구조의 연관성을 분석해본다.

1. 교통카드의 데이터 구성

지하철을 그룹별로 분류하기 위해서는 지하철역의 정보들이 필요하다. 수도권 지하철역의 개수는 2005년도 기준 357개의 역이 있고, 그 중 환승역은 54개역이다. 이들을 교통카드의 대용량 데이터를 이용하여 각 역의 정보들을 포함하는 자료구조를 형성한다. 각 역을 정점으로 나타내고 정점은 지하철역ID, 지하철역명, 지하철 노선 번호, 위도, 경도, group번호 등의 정보를 담고 있으며, 이 정점들을 link로 나타내면 하나의 그래프로 나타낼 수 있다. 만들어진 그래프에 대해 지하철 교통카드 트랜잭션을 이용하여 승객수를 알아볼 수 있다. 트랜잭션은 교통카드에 관한 정보들을 가지고 있는데 교통카드번호, 승차일시, 승차역ID, 하차역ID, 환승횟수 등의 속성을 포함하고 있다.

아래의 [표 1]은 속성과 그에 해당하는 트랜잭션의 예제를 나타낸 것이다 [5]. 속성에서는 교통카드 트랜잭션에 포함되어 있는 정보들을 나타낸 것이

고, 트랜잭션의 예제에서는 그 속성에 해당하는 정보들을 모아놓은 것이다. 속성 중에서도 승차일시와 하차일시에 해당하는 트랜잭션의 정보들을 보면 지하철 이용객의 승, 하차 시간을 알 수 있어서 시간대별 분석을 할 수 있으며, 승차 정류장 및 하차 정류장의 정보로 인하여 역의 정보를 얻어내어 승, 하차의 승객 흐름을 분석할 수 있다. [표 1]에서 첫 번째 행의 값은 교통 카드 트랜잭션의 속성 24가지를 보여주고 있다. 나머지는 승객이 사용한 교통카드의 트랜잭션들을 나타내고 있다. 트랜잭션 예제 1과 2는 한 승객이 아침에 독산역에서 지하철을 승차하여 신길역에 하차한 후에 약 3시간의 업무를 보거나 또는 머문 후에 다시 신길역에서 승차하고 독산역에서 하차한 트랜잭션을 보여주고 있다. 트랜잭션 3과 4는 어떤 승객이 아침에 9시 6분 55초에 연신내역에서 승차하여 9시 44분 19초에 압구정역에 하차하여 약 11시간 후인 오후 8시 55분 49초에 압구정에서 승차하여 처음 출발역이 아닌 녹번역에 하차한 것을 보여주고 있다.

[표 1] 교통카드 트랜잭션의 속성과 예제[5]

속성 (attribute)	카드번호, 승차일시, 트랜잭션ID, 교통수단CD, 환승횟수, 버스노선ID, 버스노선명, 교통사업자ID, 교통사업자명, 차량ID, 차량등록번호, 사용자구분코드, 사용자구분명, 운행출발일시, 승차정류장ID, 승차정류장명, 하차일시, 하차정류장ID, 하차정류장명, 이용객수_다인승, 승차금액, 승차위반금액, 하차금액, 하차위반금액
트랜잭션 예제 1	10329,20050624093758,015,200,0,,211100000,한국철도공사,,01,일반,,1714,독산,20050624095917,1032,신길,1,800,0,0,0
트랜잭션 예제 2	10329,20050624132650,016,200,0,,211100000,한국철도공사,,01,일반,,1032,신길,20050624134640,1714,독산,1,800,0,0,0
트랜잭션 예제 3	10418,20050624090655,005,200,1,,211000000,서울지하철공사,,01,일반,,0311,연신내,20050624094419,0326,압구정,1,300,0,200,0
트랜잭션 예제 4	10418,20050624205549,006,200,0,,211000000,서울지하철공사,,01,일반,,0326,압구정,20050624212748,0313,녹번,1,800,0,100,0

2. 분류 방법

수도권 지하철의 모든 역들을 그룹으로 나누기 위해서 가장 중요한 점은 분류방법이다. 어떤 분류를 통한 분석이냐에 따라 결과가 달라질 수 있기 때문이다. 따라서 분류하기 위한 방법으로 본 논문에서는 크게 세 가지 방법으로 나누었다.

첫 번째 방법으로 서울의 한강을 기준으로 한강의 남단과 한강의 북으로 나누어, 한강의 남쪽은 강남그룹, 북쪽을 강북그룹으로 칭하여 2그룹으로 나눈 후, 다시 강남그룹과 강북그룹에서 각각 두 개의 그룹으로 나누어 4개의 그룹으로 분류 후에 분석한다.

두 번째 방법으로 초기 중심점을 랜덤으로 정하는 K-means 알고리즘을 이용하여 4개의 그룹으로 나누어 분석한다. 그리고 2단계로, 초기 중심점을 랜덤이 아닌 다른 방법으로 설정하여 분석한 후, 초기 중심점이 랜덤일 때의 승객 흐름과 다른 방법으로 승객 흐름을 분석한 결과를 비교분석한다.

마지막으로 MST(Maximum flow Spanning Tree)기준으로 시간대별 승객 흐름이 많은 역을 분석하여 주요 역을 선택하고 이를 토대로 아침, 점심, 저녁시간대의 승객 흐름을 분석 한다. 각 분류 방법을 통해 그룹별 승, 하차가 활발히 나타나는 지역을 알아볼 수 있다.

2.1 한강을 기준으로 그룹 분류

예로부터 한강은 한반도의 중심에 자리 잡고 있어, 식량과 물을 쉽게 구할 수 있고 교통의 중심지로써 많은 사람들이 모여 살았다. 현대의 지금까지도 한강을 중심으로 서울은 수도권으로 한국의 중심에서 중요한 역할을

하고 있다. 이에 따라 서울은 한강을 중심으로 지역을 나누는데 강을 중심으로 강의 남쪽을 강남, 강의 북쪽을 강북이라 칭한다. 수도권 지하철을 지리학적으로 서울의 핵심인 한강을 중심으로 한강의 남쪽에 위치하는 역들을 강남그룹, 한강의 북쪽에 위치하는 역들을 강북 그룹이라고 지정한다. 또한 강남그룹과 강북그룹으로 분류한 2개의 그룹들을 다시 각각 2개의 그룹으로 나누어 총 4개의 그룹으로 분류한다.

2.1.1 분류 단계

다음은 한강을 기준으로 수도권 지하철역을 분류하기 위한 과정이다.

- 1) 1단계: 지하철역을 한강을 중심으로 강남과 강북으로 그룹을 나눈다. 지하철역의 위치상 한강의 남쪽에 위치하면 강남그룹, 한강의 북쪽에 위치하면 강북그룹으로 나누어 모든 지하철역을 2개의 그룹으로 분류한다.
- 2) 2단계: 1단계에서 나눈 강남그룹과 강북그룹을 다시 4개의 그룹으로 나누기 위해 강남과 강북그룹에서 각각 승, 하차가 활발히 이루어지는 2개의 역을 임의로 선정한다. 4개의 역을 선정하여 Floyd Shortest Path 알고리즘 [14]을 통해 모든 역을 4개의 그룹으로 분류한다.

2.1.2 Floyd Shortest Path 알고리즘

최단경로를 구하기 위한 Floyd Shortest Path 알고리즘은 모든 정점에서 출발해서 출발한 정점을 제외한 모든 정점을 도착점으로 하는 최단거리를 구하는 방법이다. Floyd Shortest Path 알고리즘은 2차원 배열을 이용하여 3중 반복을 하는 루프로 구성한다.

```

void floyd (int n const number W[][], number D[][])
{
    index i, j, k;

    D=W;
    for(k = 1; k <= n; k++)
        for(i = 1; i <= n; i++)
            for(j = 1; j <= n; j++)
                D[i][j]=minimum(D[i][j], D[i][k]+D[k][j]);
}

```

[그림 2] Floyd Shortest Path Algorithm

[그림 2]는 Floyd Shortest Path 기본 알고리즘을 나타내고 있다[14]. 그래프의 정점의 개수는 n 이고 그래프는 2차원 배열 $W[i][j]$ 로 나타낸다. $D[i][j]$ 는 정점 i 에서 정점 j 로 가는 최단 경로의 길이가 된다. 먼저 D^{k-1} 까지는 완벽한 최단 거리가 구해져서 있다고 가정하자. 일반적으로 k 번째 정점이 추가로 고려되는 상황을 생각하여 보자. $D^0 = W$ 로 초기화 한다고 가정하면, 0부터 k 까지의 정점만을 사용하여 정점 i 에서 정점 j 로 가는 최단 경로는 다음의 2가지의 경우로 나누어서 생각할 수 있다. 첫 번째는 정점 k 거쳐서 가지 않는 경우이다. 이 경우에는 k 보다 큰 정점은 통과하지 않으므로 이 경우 최단 거리는 $D^{k-1}[i][j]$ 가 된다. 정점 k 를 통과하는 경우에는 i 에서 k 까지의 최단거리 $D^{k-1}[i][k]$ 에다가 k 에서 j 까지의 최단 거리인 $D^{k-1}[k][j]$ 를 더한 값이 될 것이다[16]. [표 2]에서는 $D = W$ 로 초기화 되고, 인덱스 k 가 1에서 n 까지 변화하여 D^k 가 D 로 사용되고 있다.

2.1.3 Floyd Shortest Path 알고리즘을 통한 데이터 분석

Floyd Shortest Path 알고리즘을 통해 4개의 중심지로 지정한 역과 가까

운 역들을 찾아내어 4개의 그룹을 만들 수 있다. 다음 [그림 3]은 실제 Floyd Shortest Path 알고리즘을 이용하여 그룹으로 분류하는 과정의 예이다.

```
for n개의 모든 각각의 역 x에 대해 {
  if 강북그룹 {
    if ([Hop[x][종각] < Hop[x][노원])
      x = 종각그룹;
    else
      x = 노원그룹;
  }else 강남그룹 {
    if (Hop[x][강남] < Hop[x][영등포])
      x = 강남그룹;
    else
      x = 영등포그룹;
  }
}
```

[그림 3] 4개의 그룹으로 분류하기 위한 알고리즘의 예제

[그림 3]의 예제는 수도권 지하철의 각각의 역 n개중 x를 강북그룹과 강남그룹 중 또다시 4개의 역 종각역, 노원역, 강남역, 영등포역 그룹으로 나누기 위한 알고리즘을 나타낸 것이다. n개의 모든 역들을 위의 분류 단계에서 보았을 때 강북그룹과 강남그룹으로 나누었던 단계를 거쳐 4그룹으로 나눈다. 예를 들어 시청역을 4그룹중 하나의 그룹으로 배정한다고 하자. 시청은 이미 강북그룹에 속해 있으므로, 강북그룹의 종각역과 노원역과의 비교로 더 가까운 역에 배치된다. 시청은 종각 그룹에 더 가까우므로, 종각그룹이 된다. 또 다른 역인 서초역을 예로 들어 보자면 서초역은 강남그룹에

속해있으므로 강남역과 영등포역과의 비교로 그룹이 결정된다. 두 역과 비교 시 강남역과 더 가까우므로 서초역은 강남그룹이 된다. 이렇게 모든 역들을 4개의 그룹으로 분류할 수 있다.

2.2 K-means 알고리즘을 이용하여 그룹으로 분류

위의 그룹을 분류하는 방법에서는 한강을 기준으로 강남과 강북을 나누어 임의의 4개의 역을 지정하여 모든 역들을 그중 가장 가까운 역에 배치하는 분류 방법 이었다. 초기의 역을 지역적인 위치로 인한 임의의 선택하는 대신에 알고리즘을 이용하여 선택하고 주요 역을 비교해 본다.

K-means 알고리즘은 n개의 초기 중심점을 선택하면 모든 역들이 각 중심점을 바탕으로 가장 가까운 역으로 배치가 된다. 모든 역을 배치하고 나면, 각각의 그룹의 중심점을 다시 계산하여 재배치하는 알고리즘이다. 이러한 K-means 알고리즘을 이용하면 초기의 중심점이 재계산되어 계속 바뀌기 때문에 위의 지리학적으로 주요 역을 선택하여 분류한 방법과 알고리즘을 이용하여 분류한 주요역과 비교해볼 수 있다.

2.2.1 분류단계

K-means 알고리즘은 보통 초기 중심점을 어떻게 선택하느냐가 가장 중요한 논쟁이 된다. 초기 값이 무엇이냐에 따라 그 결과도 다를 수 있기 때문이다. 본 논문에서는 초기 중심점을 두 가지 방법으로 선택하여 분류한다. 그 두 가지 방법은 다음과 같다.

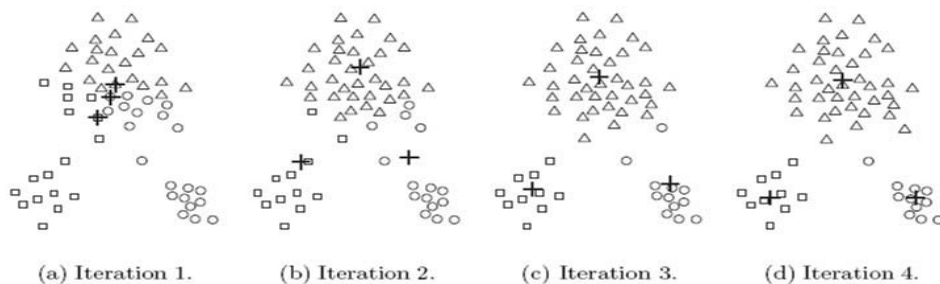
- 1) 방법 1: 4개의 초기 중심점을 임의로(random selection) 설정하여 4개

의 그룹으로 분류한다.

2) 방법 2: 초기 중심점을 설정하는 방법을 다양화 하는 방법 중 하나인 지하철 노선도 상에서 가장 끝 부분에 위치한 4개의 지하철역을 지정하여 분류한다.

2.2.2 K-means 알고리즘

K-means 알고리즘은 데이터마이닝의 군집분류 방법 중 하나로써, n개의 초기 중심점을 생성하여 군집을 형성한다. 형성된 군집을 다시 계산하여 군집의 중심점을 구하고 다시 모든 점들을 재배치한다. 이 과정을 초기중심점이 바뀌지 않을 때까지 반복하여 분류하는 방법이다.



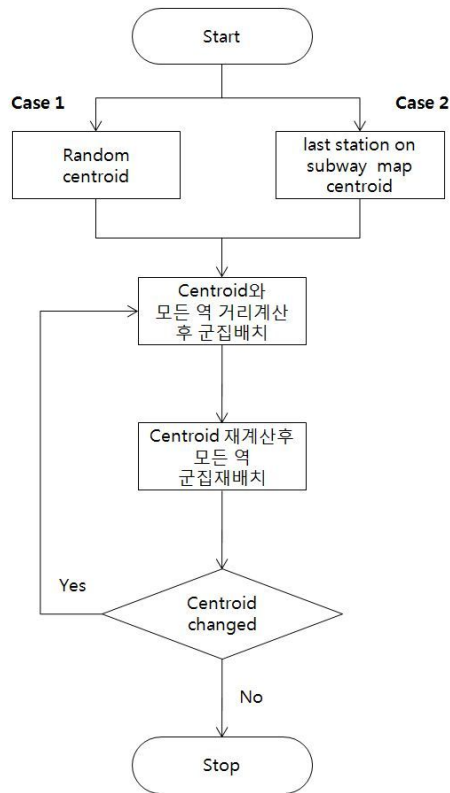
[그림 4] K-means 알고리즘의 예

[그림 4]는 세 개의 군집을 찾기 위한 K-means 알고리즘의 과정의 한 예이다[15]. 첫 번째 단계에서 초기의 중심점들은 모두 큰 점들의 집합에 속해 있다. 각 군집들의 평균을 구하여 중심점을 갱신한다. 2단계에서 보면 1단계와는 달리 큰 군집에서 모여 있던 세 개의 중심점들이 각 소군집에 배치되어 있는 것을 볼 수 있다. 3단계에서는 각 군집에서 중심점이 중심으로 이동한 것을 볼 수 있다. 마지막으로 군집의 중심점이 변하지 않을 때까지 반

복하는 것을 볼 수 있다. 이렇게 군집의 중심점을 구하고 분류하는 알고리즘이라고 할 수 있다.

2.2.3 K-means 알고리즘을 이용한 데이터 분석

위에서 살펴보았던 K-means 알고리즘을 이용하여 수도권 지하철역을 4개의 그룹으로 분류하는 과정을 이해를 돕기 위해 아래의 [그림 5]로 나타내보았다.



[그림 5] K-means 알고리즘을 이용하여 역 분류를 위한 순서도

[그림 5]의 case 1과 case 2는 초기 중심점을 어떻게 선택할지에 대한 결정이다. 경우 1은 초기 중심점을 랜덤으로 설정하기 때문에 임의의 4개의 역이 선택되어 진다. 경우 2는 지하철 노선도 상에서 양쪽 끝에 위치한 역을 초기 중심점으로 설정하는 방법이다. 초기 중심점을 정하고 나면, 중심점 위주로 모든 역들은 위치상 가장 가까운 곳으로 배치되어 군집을 형성한다. 모든 역이 4개중 한 군집으로 배치가 되면 다시 4개의 군집의 중심점을 재계산 한다. 재계산된 중심점을 위주로 다시 모든 역에 대한 군집이 재배치

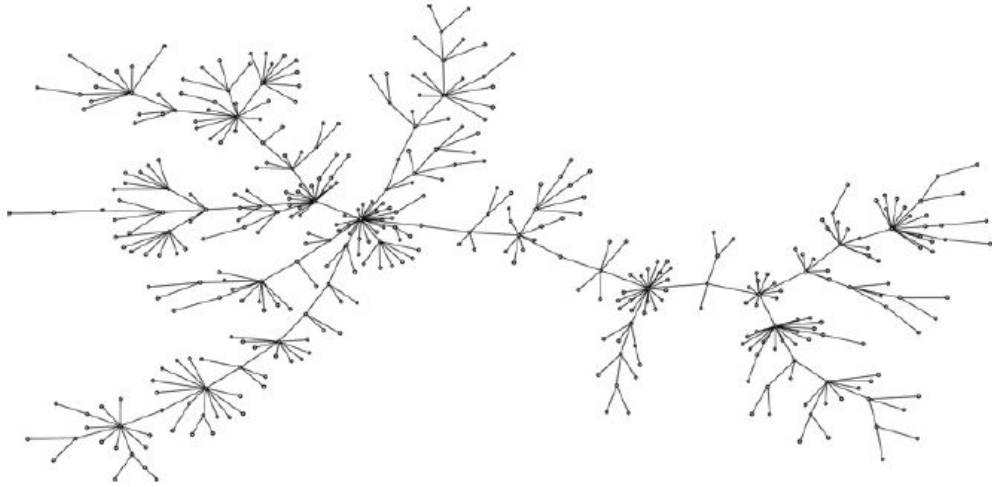
가 되고, 이 과정은 중심점이 바뀌지 않을 때까지 반복된다. 그리고 마지막으로 분류된 군집이 최종 분류가 된다.

2.3 MST 기준 시간대별 그룹으로 분류

위의 두 가지 방법에서는 한강을 기준으로 강북과 강남 그룹으로 나누어 비교해보고 다시 4개의 임의의 역을 지정하여 모든 역을 4그룹으로 나누기 위하여 최단경로 알고리즘을 이용하여 그룹을 분류하는 방법 이였고, 또 K-means 알고리즘을 이용하여 초기 중심점을 랜덤으로 설정하여 분류해보고, 지하철 노선도 상에서의 양쪽 끝의 역을 선택하는 방법 이였다. 이는 모든 역을 임의로 선택하거나 랜덤으로 설정하여 알고리즘을 통하여 역을 분류하는 방법이다. 이번에는 임의로 선택하거나 알고리즘을 이용하여 랜덤으로 설정하는 방법 대신 MST(Maximum flow Spanning Tree)를 이용하여 시간대별 승객 흐름이 가장 많은 역 10개를 선정하여 그 안에서 주요 역을 설정하고 그룹 역을 분류하는 방법이다. 위의 두 방법과는 다르게 아침, 점심, 저녁으로 구분하여 승객의 흐름을 시간대별로 분석해 볼 수 있다.

2.3.1 MST(Maximum flow Spanning Tree)를 이용한 분석

[그림 6]은 참고문헌 [13]에서 참조한 MST 그래프이다. 이 그래프는 두 지하철 사이의 승객 흐름이 최대가 되는 edge를 차례로 찾아서 spanning tree를 만든 결과이다. 이 Tree에 기초하여 각 클러스터의 중심이 되는 지하철역을 선택하였다.



[그림 6] 서울 수도권 지하철 시스템에서 승객 흐름에 대한 Maximum flow Spanning Tree(MST)[13]

2.3.2 분류단계

다음은 MST를 이용하여 승객흐름을 분석하기 위한 그룹 분류의 과정이다.

1) 1단계: MST 기준으로 시간대별로 morning, day, evening의 기준으로 가장 많은 승객수의 역 10개를 선정한 결과를 이용하여, 10개의 역중 가장 많은 역을 가지고 있는 그룹 8개의 역을 선택한다. 또한 8개의 그룹을 다시 4개의 그룹으로 나누기 위해 이미 분류되어진 8개의 역을 다시 가장 많은 역을 보유하고 있는 그룹 4개를 선택하여 분류한다. 단, 가장 많은 그룹 역을 가지고 있는 역 4개중 가까운 거리에 있는 역은 둘 중 하나를 선택하고 그 다음으로 많은 역을 선택한다.

2) 2단계: 전체 하루 동안의 주요 역을 구하기 위해 모든 시간대의 역들을 통합하여 가장 많은 역순으로 설정한다. 최초에 선정되었던 전체 30개의 역

중 그룹역의 수가 10개 미만인 역을 제외하고 다시 분류한다. 그리고 그룹역의 개수가 많은 순으로 12개의 역을 선정하여 8개의 역, 4개의 역으로 나눌 수 있을 때까지 반복하여 최종 4개의 역을 선택한다.

2.3.3 MST(Maximum flow Spanning Tree)를 이용한 지하철 역 분류 방법

[표 2] 하루 동안의 시간대별 승객 흐름으로 분류한 역

rank	morning	day	evening
1	강남	고속터미널	고속터미널
2	종각	부평	부평
3	수원	천호	강남
4	청담	수원	천호
5	연신내	강남	종각
6	선릉	노원	수원
7	광화문	영등포	노원
8	신림	연신내	영등포
9	양재	종각	신촌
10	마포구청	잠실	서현

[표 2]는 MST기준으로 하루 동안의 시간대별 승객의 흐름이 가장 활발히 일어나고 있는 역을 순위별로 나타난 표이다. 각각의 시간대에는 10개의 역이 선정되었다. [표 2]를 보면 낮 시간대와 저녁시간대의 선택된 역이 비슷한 것으로 보아 승객흐름도 비슷한 것으로 보인다. 이 중의 역에서 아침, 점심, 저녁의 시간대의 역을 각각 그룹별 개수가 많은 역 8개가 선택되어지고 다음으로 4개의 역이 선택되어 진다.

각각의 시간대별 나누어진 아침, 점심, 저녁의 그룹은 시간대별로 승객흐

를 구하기 위해서는 2005년 6월 24일의 교통카드 정보를 이용하여 base time 설정을 아침시간은 10시 이전, 저녁시간은 17시 즉 오후 5시 이후로 설정하고 중간시간에 해당되는 시간은 점심으로 설정한다. 설정된 시간은 교통카드 트랜잭션의 속성에 의해 나뉘지게 되는데, 교통카드의 속성에는 많은 정보들이 있지만 그중에서도 승차일시와 하차 일시의 정보로 인하여 아침, 점심, 저녁 중 어느 시간대에 승차하고 하차하였는지 구분이 가능하다.

IV. 실험결과 및 분석

1절에서는 실험에서 사용한 컴퓨터의 환경과 데이터를 설명하고 2절에서는 3장에서 제시한 그룹을 분류하기 위한 방법 세 가지로 인한 승객 흐름을 분석하고 설명한다.

1. 실험환경

실험을 하기 위해 사용되어진 데이터는 2005년 6월 24일의 교통카드 트랜잭션의 정보를 사용했으며, 지하철 노선도는 2005년 기준으로 지하철역의 개수는 357개의 역으로 구성되어 있고 이중에서 환승역은 54개역이다.

하루 동안의 기록된 전체 트랜잭션의 개수는 10,667,519이며, 이중에서 정보 불분명 건수가 679,485건으로 분석대상 통행 기록 개수는 총 9,988,034개이다. 이 개수는 버스 이용건수와 지하철 이용건수가 합해진 값이므로, 이중 에러 트랜잭션을 제외한 지하철 이용건수 총 4,909,153개의 데이터를 가지고 실험했다. 실험에 사용된 운영체제는 Microsoft Window 7 32비트 이고, 자세한 컴퓨터 하드웨어 사양은 3.0GHz Intel Core(TM)2 Duo CPU와 2GB의 메모리를 사용했다. 그리고 사용된 언어는 Microsoft Visual Studio 2008 C++ 언어이다.

2. 실험결과 및 분석

이 절에서는 위에서 제시한 세 가지 방법들의 실험결과를 나타내었다. 각 절에서는 각 방법에 대한 실험 결과와 그에 따른 분석을 나타낸다.

2.1 한강을 기준으로 그룹으로 분류

아래 [표 3]은 1단계로 4개의 그룹으로 분류하기 전 2개의 그룹으로 분류하였을 경우의 그룹개수와 2단계인 4개의 그룹으로 분류했을 때의 그룹개수를 나타내고 있다. 각 단계의 구분은 한강을 기준으로 한강의 남단에 있는 역을 강남그룹이라 하고 한강의 북쪽에 있는 역을江北그룹이라 칭했다. 1단계에서는 지도상에서 나타낸 역의 위치를 눈으로 구분하여江北그룹은 0, 강남그룹은 1로 표시하여 구분하였다. 2단계에서는 1단계에서 나뉜江北그룹과 강남그룹에서 중심역이라고 생각되어지는 역을 임의로 선택하여 4개의 그룹으로 나누었고, 2단계에서는 1단계에서와 달리 Floyd Shortest Path 알고리즘을 이용하여 모든 역들을 각 중심역과 가장 가까운 곳에 위치하는 곳에 배치하였다.

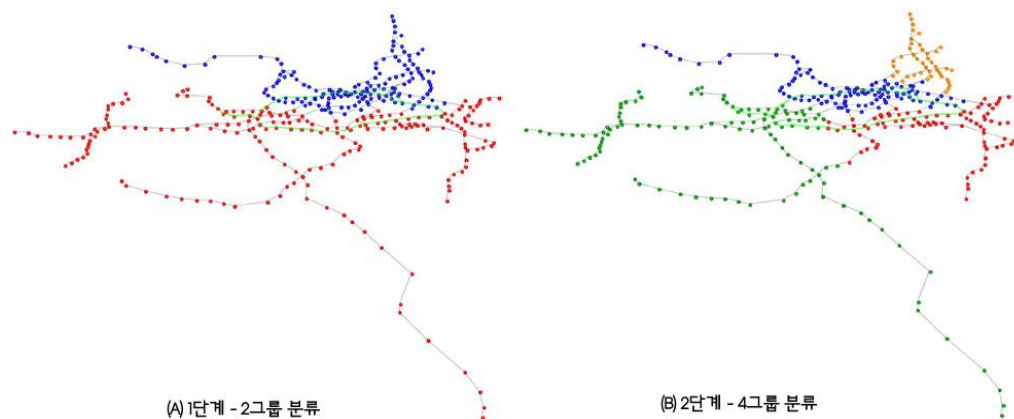
[표 3] 1, 2단계의 각 그룹 역 개수

1단계		2단계			
江北그룹	강남그룹	종각역그룹	노원역그룹	강남역 그룹	영등포역 그룹
150	207	108	42	87	120

[표 3]에서 보는바와 같이 강남그룹이 57개의 역이 더 많은 것을 볼 수 있다. 2단계에서는 영등포역 그룹이 가장 많고 그 다음 종각역 그룹으로 나타

났다. 4그룹 중 종각역과 노원역은 강북에 해당하며 강남역과 영등포는 강남그룹에 해당하는데 강북그룹에서는 종각역 부근으로 역이 많이 몰리는 것을 볼 수 있으며 강남그룹에서는 영등포역중심으로 많은 역이 있다는 것을 알 수 있다.

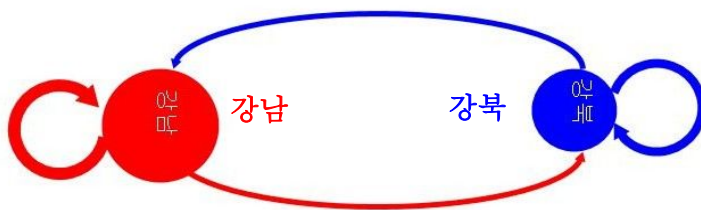
[그림 7]은 실제 지하철역의 위치를 적용해 노선도로 나타낸 것이다. 왼쪽 그림은 1단계의 역 분류인데 빨간색은 강남그룹, 파란색은 강북그룹으로써 보기에 빨간색 역의 그룹인 강남그룹이 더 많음을 알 수 있다. 오른쪽은 2단계 분류로써 앞에서 분류한 2개의 그룹에서 4개의 그룹으로 분류한 것이다. 파란색은 종각역, 노란색은 노원역, 빨간색은 강남역, 초록색은 영등포역 그룹을 나타낸다. 4개의 그룹에서는 파란색과 초록색의 영등포그룹과 종각 그룹역이 많은 것을 볼 수 있다.



[그림 7] 지하철 노선도 상에서 역 분포

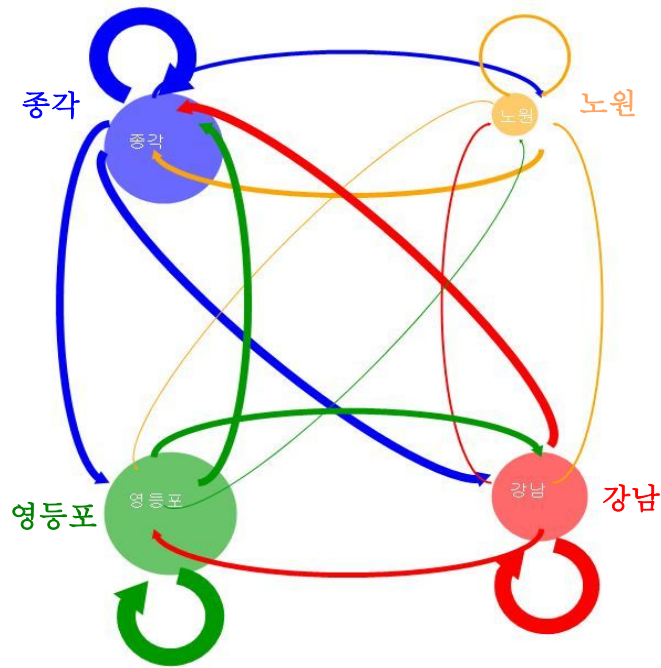
[그림 8]은 1단계로 의 승객 흐름으로써 출발→도착 그룹유형으로 승객수를 MFC를 이용해 승객수를 대입하여 시각화한 것이다. 기본 라인으로 선의 폭은 5로 지정해줬으며, 이 기본 라인은 그룹 중에서 가장 승객수가 낮은

그룹이 기본 라인의 기준이 된다. 다시 말하면 가장 적은 승객수의 그룹이 기준이 되어 다른 그룹의 승객 수에 따라 굵기는 비례하게 된다. 또한 그룹을 나타내고 있는 정점은 실제 역 개수에서 1/2배 축소하여 크기를 나타낸 것으로, [그림 8]을 분석하여 볼 때 출발해서 도착 시 강북→강북 이동그룹과, 강남→강남 이동그룹이 다른 그룹에 비해 승객수가 많은 것을 볼 수 있다. 이는 대부분의 승객들이 같은 범위 안에서의 이동이 더 많다는 것을 알 수 있다.



[그림 8] 1단계 그룹 승객 흐름도

다음 [그림 9]는 2단계를 나타낸 것으로써 마찬가지로 MFC를 이용하여 승객수를 대입하여 시각화한 것이다. 2단계에서는 기본 라인의 선의 폭은 2로 지정해주었고, 그룹을 나타내는 정점의 크기는 각 그룹역의 실제 역 개수로써 나타내주었다. 2단계 역시 노원그룹을 제외한 모든 그룹이 같은 그룹에서의 출발→도착이 승객수가 가장 많은 것을 볼 수 있다. 노원그룹은 노원그룹보다 종각그룹에서의 도착 승객수가 더 많아서 이런 결과는 지역적으로 강북그룹에서 종각역이 주요역임을 알 수 있다. 1단계에서도 보았듯이 2단계 역시 같은 지역 내에서의 승, 하차가 활발히 나타나는 것을 더 자세히 분석할 수 있다. 그러나 이러한 결과는 같은 그룹역의 개수가 얼마나 많이 차이가 나느냐에 따라 결과가 달라질 수 있기 때문에 이러한 점을 유의해야 한다.



[그림 9] 2단계 그룹 승객 흐름도

다음 [표 4]는 [그림 8]과 [그림 9]의 승객수를 표로 나타낸 것이다. 위에서 본 것과 같이 1단계에서는 강남→강남 이동그룹의 승객수가 가장 많으며 그 다음 강북→강북 이동그룹이 많은 것으로 보여 진다. 2단계에서는 종각→종각 이동그룹, 강남→강남 이동그룹, 영등포→영등포 이동그룹을 보면 위의 [그림 9]와 같이 가장 많은 승객수를 나타내고 있으며 이 세 그룹들 간의 이동이 그다음으로 많은 것으로 나타났다.

[표 4] 1, 2단계 지하철 승객 수

1단계	출발 \ 도착	강북그룹		강남그룹		합계
	강북그룹	1,349,037		820,364		2,169,401
	강남그룹	821,801		1,917,951		2,739,752
	합계	2,170,838		2,738,315		4,909,153
2단계	출발 \ 도착	종각그룹	노원그룹	강남역그룹	영등포그룹	합계
	종각그룹	762,554	198,058	348,913	312,250	1,621,775
	노원그룹	223,622	164,803	112,639	46,562	547,626
	강남역그룹	349,074	92,800	673,049	233,257	1,348,180
	영등포그룹	337,092	42,835	257,915	753,730	1,391,572
	합계	1,672,342	498,496	1,392,516	1,345,799	4,909,153

2.2 K-means 알고리즘을 이용한 그룹분류

다음 [표 5]는 K-means 알고리즘을 이용하여 그룹을 분류한 결과이다. start는 처음 선택되어진 4개의 역을 나타내고, last는 마지막으로 분류되어진 4개의 역의 개수를 나타낸다. 그리고 last station의 그룹개수는 마지막으로 분류되어진 그룹의 개수를 나타내고 있다.

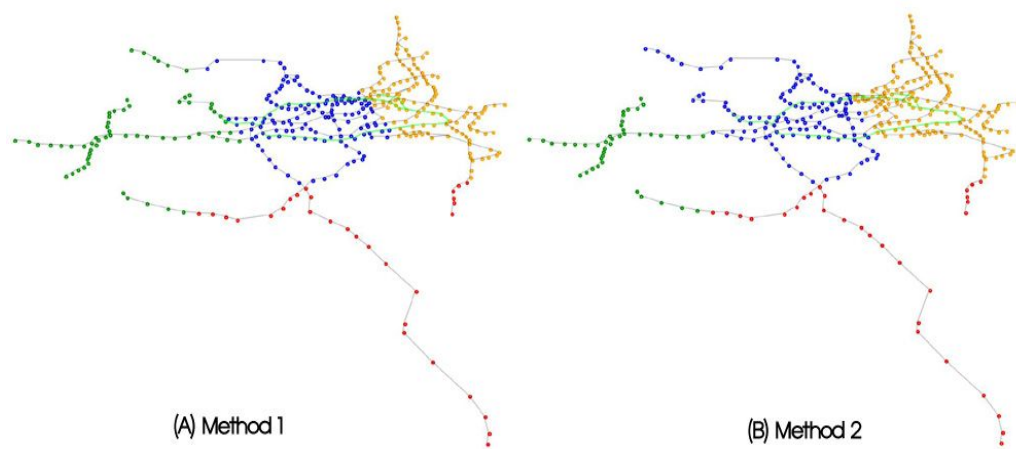
[표 5] K-means 알고리즘으로 분류한 단계별 그룹 개수

1단계			2단계		
start	last	last station 그룹개수	start	last	last station 그룹개수
문래	송내	63	인천	간석오거리	42
응암	군자	129	의정부북부	장한평	161
증산	용산	135	대화	당산	124
수서	세류	30	오리	세류	30

[표 5]의 1단계에서는 K-means 알고리즘의 초기 중심점을 랜덤으로 설정하여 분석한 결과이다. 처음 시작된 중심점이 K-means 알고리즘을 통해 재배치된 후, 마지막으로 분류된 역으로 송내역, 군자역, 용산역, 세류역으로 결과가 나타났다. 지하철 노선도에서 살펴보면 처음보다는 배치가 균등하게 분류된 것으로 보이나 각 그룹역의 개수를 볼 때 역들이 군자와 용산에 치우쳐져 있다는 것을 알 수 있다.

K-means 알고리즘의 가장 큰 단점은 초기 중심점을 어떻게 결정하느냐에 따라 결과 값이 달라질 수 있다는 점인데, 초기중심점을 잘 정할수록 결과 값이 균등하게 배치된다. 이를 보완하기 위해 2단계에서는 초기중심점을 랜덤으로 설정하지 않고 지하철 노선도 상에 있는 각 끝부분에 해당하는 역들을 선택해보았다. 그 결과 간석오거리역, 장한평역, 당산역, 세류역이 마지막 분류역의 중심점이 되었으며, [표 5]의 각 그룹역의 개수를 살펴볼 때 1단계에서와 마찬가지로 장한평역과 당산역 그룹으로 많이 몰려있는 것을 볼 수 있다. 이는 K-means 알고리즘의 초기 중심점 선택이 쉬운 문제가 아니라는 것을 느낄 수 있다. 따라서 K-means 알고리즘의 성능을 최대한으로 활용하기 위한 초기중심점 선택의 연구가 필요하다.

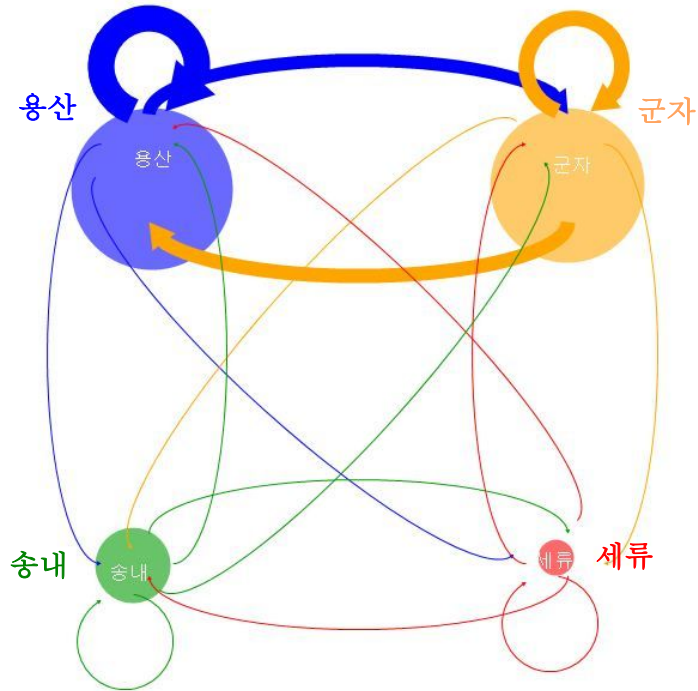
아래 [그림 10]은 실제의 지하철역의 위치를 적용하여 지하철역을 노선도로 표현하였으며, (A) method1(1단계)의 파란색으로 표시된 그룹은 용산역 그룹, 노란색은 군자역그룹, 초록색은 송내역그룹, 빨간색은 세류역 그룹을 나타낸다. 그리고 (B) method2(2단계)의 파란색은 장한평역 그룹, 노란색은 당산역 그룹, 초록색은 간석오거리역 그룹, 빨간색은 세류역그룹이다. 아래의 그림을 통해 보자면 위의 표에선 알 수 없던 역들의 배치를 알 수 있어 시각적으로 확인하기에 편리하다.



[그림 10] k-means 알고리즘을 이용하여 분류한 역 분포

아래의 [그림 11]과 [그림 12]는 위의 그룹 분류를 통한 그룹별 승객의 흐름을 시각화 한 것이다. 위의 분류방법에서와 마찬가지로 MFC로 표현하였으나, 각 그룹별 승, 하차 승객수의 큰 차이 때문에 각 그룹 중 최소 값을 대신하여 모든 승객 수에 대한 평균값을 구하여, 평균값 이상의 값들에 대해서 기본 라인 값에 비례하게 나타내어졌다. 기본 라인 값은 5로 지정해주었으며 평균값 이하의 값들은 모두 1로 표현해주었고, 마찬가지로 그룹을 나타내는 정점의 크기는 실제 그룹의 역 개수로 표현하였다. [그림 11]은

K-means 알고리즘을 이용한 초기중심점을 랜덤으로 선택했을 때의 승객 흐름도이다. 용산역→용산역 이동그룹의 승객수가 가장 많은 것으로 보이며 이는 위에서 보았던 것처럼 용산역과 군자역의 그룹분류 개수가 많기 때문이기도 하다.

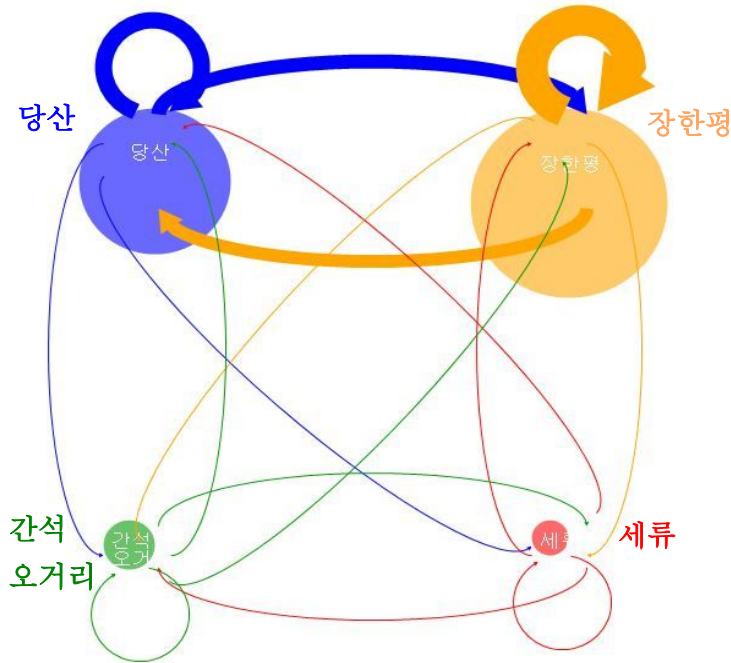


[그림 11] random 중심점의 승객 흐름도

[그림 12]는 [그림 11]과 같이 K-means 알고리즘을 이용하여 초기 중심점을 랜덤으로 선택하지 않고 지하철노선도 상에서 양쪽 끝에 위치한 역 4개를 선택하여 분류한 승객 흐름도이다. 위에서와 마찬가지로 두 그룹에서 눈에 띄게 큰 승객수를 보이는데, 가장 많은 승객수를 보이는 것은 장한평역→장한평역 이동그룹이며 그다음 당산역→당산역 이동그룹, 당산역→장한평역 이동그룹, 장한평역→당산역 이동그룹이다.

이 분류방법 역시 그룹별 승객수의 큰 차이로 인한 평균값을 구하여 평균

값 이상의 값을 대상으로 나타내어 주었으며, 장한평역과 당산역의 그룹개수로 인한 승객수의 차이로도 보여 진다.



[그림 12] 노선도 양 끝에 위치한 중심점의 승객흐름도

다음 [표 6]은 앞서 보았던 승객 흐름도의 단계별 승객수를 나타내고 있다. 그림에서 보았던 가장 많은 승객수를 보여주었던 그룹으로 1단계에서는 용산역→용산역 이동그룹과 2단계에서는 장한평역→장한평역 이동그룹을 살펴볼 때, 다른 그룹에 비해서 큰 차이를 보이는 것을 볼 수 있다.

[표 6] K-means 알고리즘으로 분류한 단계별 승객 수

1단계	출발 \ 도착	송내역	군자역	용산역	세류역	합계
	송내역	234,691	68,101	276,746	16,648	596,186
	군자역	63,275	835,881	749,216	52,389	1,700,761
	용산역	249,125	696,277	1,368,214	82,075	2,395,691
	세류역	16,783	53,213	88,835	57,684	216,515
	합계	563,874	1,653,472	2,483,011	208,796	4,909,153
	2단계	출발 \ 도착	간석오거리	장한평역	당산역	세류역
간석오거리		174,991	62,462	116,015	11,729	365,197
장한평역		56,584	1,645,607	723,410	70,149	2,495,750
당산역		109,192	757,420	895,845	69,234	1,831,691
세류역		11,949	73,769	73,113	57,684	216,515
합계		352,716	2,539,258	1,808,383	2,08,796	4,909,153

2.3 MST(Maximum flow Spanning Tree)기준 시간대별

그룹 분류

앞서 구분했던 2가지 방법들은 어떤 기준으로 하여 역을 선택하고 알고리즘을 이용하여 중심 역을 선택하여 분류하였다. 이번 절에서는 앞에서 했던 방식과는 달리 MST(Maximum flow Spanning Tree)의 기준으로 시간대별로 나누어 대표하는 10개의 역을 찾아내어 그중 그룹수가 가장 작은 두 개의 역을 제외한 8개의 역을 대상으로 시간대별로 8개의 그룹으로 분류하였다. 아침, 점심, 저녁의 시간별 분류와 하루 전체 동안의 그룹으로 나누어 살펴보았다.

첫 번째로 8개의 그룹으로 나누어 분석 후, 다시 4개의 그룹으로 나누어 살펴보기 위해 8개로 분류한 그룹들을 그룹 역 개수가 가장 큰 순서대로 각각 시간대의 4개의 대표 역을 선택한다. 단, 그룹 역 개수가 가장 큰 4개의 역 중에서 가까운 거리의 비슷한 역이 있으면 그 다음 순으로 그룹 역 개수가 큰 역을 선택하였다. 아래 [표 7]과 [표 8]은 시간대별 주요 역을 나타낸 것이다. 세 가지로 분류된 시간대별 그룹 분류를 합하여 하루 동안의 승객 흐름을 분석할 수 있는데 이 분류는 마지막에 살펴보기로 한다.

[표 7] MST 기준 시간대별 8개의 그룹 역 개수

Morning		Day		Evening	
그룹역이름	그룹역개수	그룹역이름	그룹역개수	그룹역이름	그룹역개수
강남	27	고속터미널	53	고속터미널	55
종각	99	부평	37	부평	37
수원	36	천호	27	천호	41
청담	41	수원	34	종각	78
연신내	21	노원	38	수원	34
선릉	40	영등포	58	노원	37
신림	76	종각	74	영등포	55
마포구청	17	잠실	36	서현	20

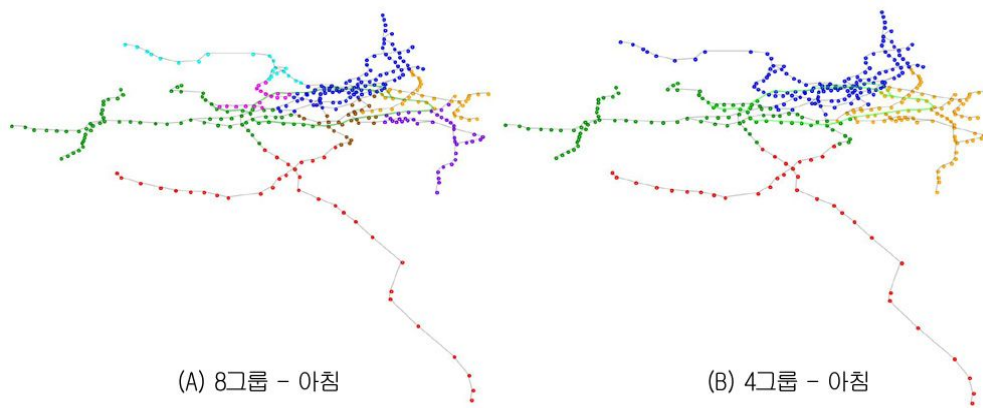
[표 7]은 시간대별 8개의 그룹 역들의 개수를 나타낸 것이다. 각 그룹별 개수를 비교해볼 때, 종각역은 어느 시간대에서도 가장 많은 그룹 역을 갖고 있는 것을 볼 수 있다. 이를 볼 때 종각은 가장 주요한 역임을 알 수 있다.

아래의 [표 8]은 8개의 그룹 역에서 가장 많은 4개의 역을 선택하여 4그룹으로 분류한 역의 개수를 나타내고 있는데, 4개의 그룹역중에서는 아침에는 종각역이 가장 많은 그룹을 가지고 있으나 낮이나 저녁때에는 그룹역이 영등포가 가장 많이 소유한 역으로 나타나고 있다. 그룹역의 개수와 승객수와는 관계가 있으므로, 그룹 역을 구분하는 방법과 개수 구분이 중요한 분석의 요인이 된다는 것을 알 수 있다.

[표 8] MST 기준 시간대별 4개의 그룹 역 개수

Morning		Day		Evening	
그룹역이름	그룹역개수	그룹역이름	그룹역개수	그룹역이름	그룹역개수
종각	135	고속터미널	114	고속터미널	59
수원	36	노원	41	천호	68
청담	91	영등포	127	종각	106
신림	95	종각	75	영등포	124

[그림 13]은 아침 시간대의 8개 그룹과 4개의 그룹분류를 노선도에 나타낸 것이다. 먼저 (A)는 8개의 그룹을 노선도 상에서 나타낸 것으로 강남그룹은 갈색, 종각그룹은 파랑색, 수원그룹은 빨간색, 청담그룹은 노란색, 연신내그룹은 하늘색, 선릉역그룹은 보라색, 신림역그룹은 초록색, 마포구청역그룹은 분홍색으로 나타내었다. 가장 많은 그룹 역은 종각역과 신림역으로써, [표 7]을 보면 자세한 그룹 역 개수를 알 수 있다. (B)는 아침시간대의 4개의 그룹 역을 노선도 상에서 색깔별로 표시하였는데, 8개의 그룹에서와 마찬가지로 종각그룹은 파랑색, 수원역그룹은 빨강색, 청담역그룹은 노란색, 신림역그룹은 초록색으로 나타내어 (A)그림과 쉽게 비교해볼 수 있다. (A)에서의 8개의 그룹들이 가장 많은 색깔의 그룹들로 가까운 곳으로 재배치된 것을 볼 수 있다.

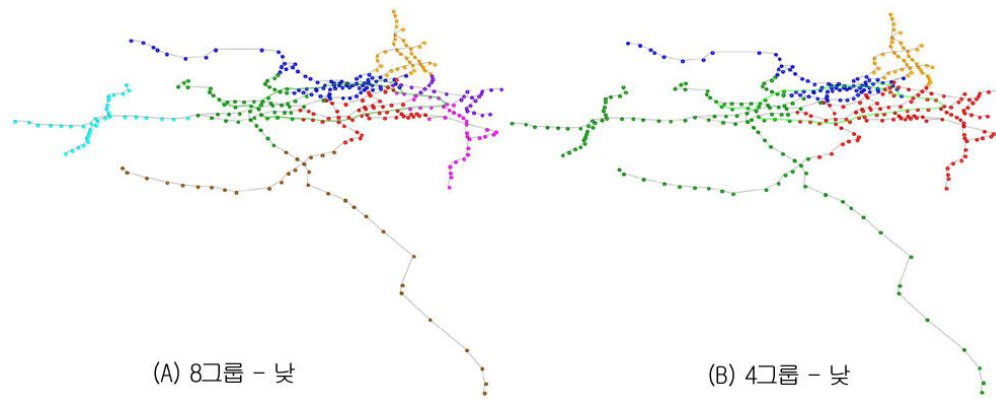


(A) 8그룹 - 아침

(B) 4그룹 - 아침

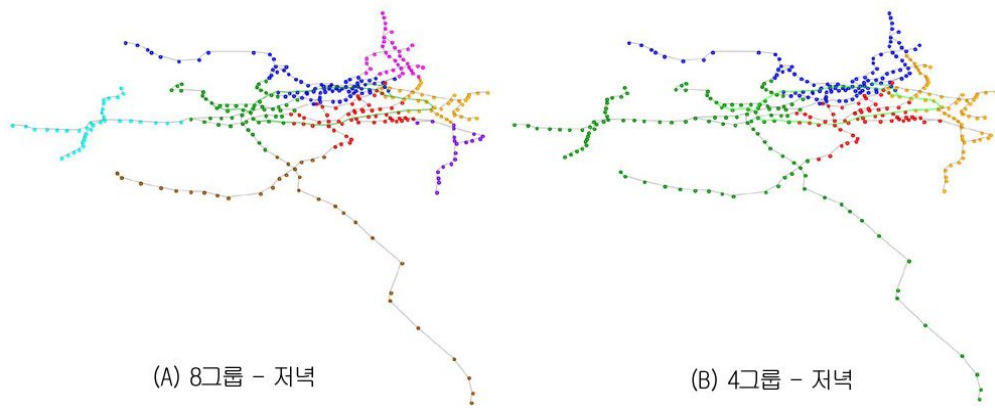
[그림 13] MST 기준 오전시간대 그룹 역 분포

[그림 14]는 MST 기준으로 낮 시간대의 주요 역들을 그룹별로 표시한 것이다. (A)그림은 8개의 그룹 역들을 색깔별로 나타낸 것으로써, 고속터미널역그룹은 빨간색, 부평역그룹은 하늘색, 천호역그룹은 보라색, 수원역그룹은 갈색, 노원역그룹은 노란색, 영등포역그룹은 초록색, 종각역그룹은 파란색, 잠실역그룹은 분홍색으로 표현했다. 이 중 종각역이 가장 많은 역을 포함하고 있으나 아침시간대의 그룹 역처럼 큰 차이가 있는 그룹 역은 없어 보인다. 다음 (B)그림은 낮 시간대의 주요역인 4개의 그룹 역을 나타내고 있으며, (A)와 마찬가지로 고속터미널역은 빨간색, 노원역그룹은 노란색, 영등포역그룹은 초록색, 종각역그룹은 파란색으로 나타내었다. 8그룹역의 분류와는 달리 초록색의 영등포그룹역과 빨간색의 고속터미널역에 속한 역이 많은 것으로 보여 진다.



[그림 14] MST 기준 낮 시간대 그룹 역 분포

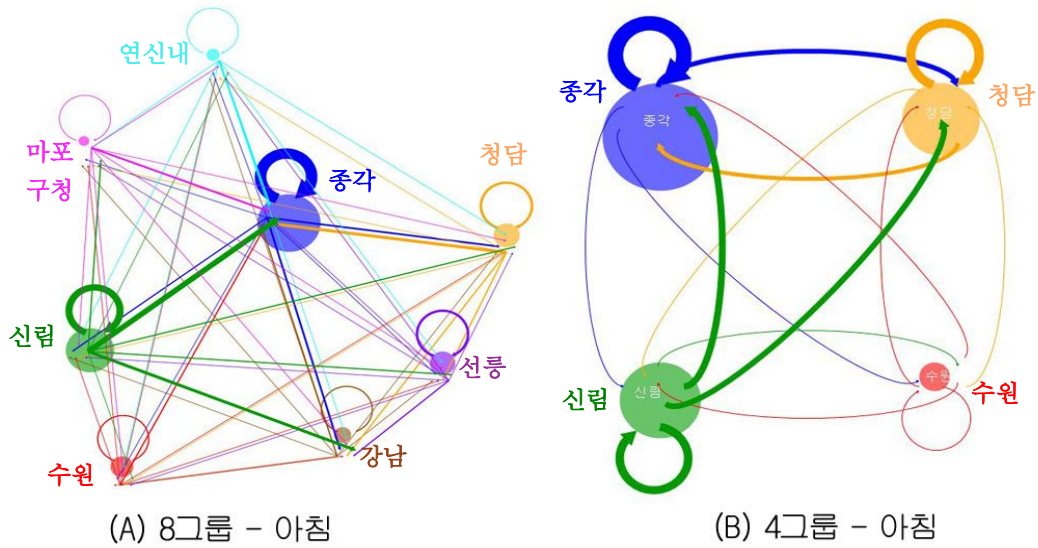
[그림 15]는 MST 기준으로 시간대별 분류한 그룹의 저녁동안의 그룹 역을 노선도 상에서 표현한 것이다. (A)는 8개의 그룹 역으로 분류한 것으로 고속터미널역그룹은 빨간색, 부평역그룹은 하늘색, 천호역그룹은 노란색, 종각역그룹은 파란색, 수원역그룹은 갈색, 노원역그룹은 분홍색, 영등포역 그룹은 초록색, 서현역그룹은 보라색으로 표현하였다. 마찬가지로 종각역의 그룹역이 가장 많으나 낮 시간대의 8개의 그룹과 같이 다른 그룹역과는 크게 차이나 보이지 않는다. 다음 (B)는 저녁시간대의 주요 역 4개의 그룹을 표현한 것으로 고속터미널역그룹은 빨간색, 천호역그룹은 노란색, 종각역그룹은 파란색, 영등포역 그룹은 초록색으로 나타내었다. 4그룹 분류에서는 영등포역그룹과 종각역그룹이 다른 그룹들에 비해 포함하고 있는 역이 많은 것을 볼 수 있다.



(A) 8그룹 - 저녁

(B) 4그룹 - 저녁

[그림 15] MST 기준 저녁 시간대 그룹 역 분포



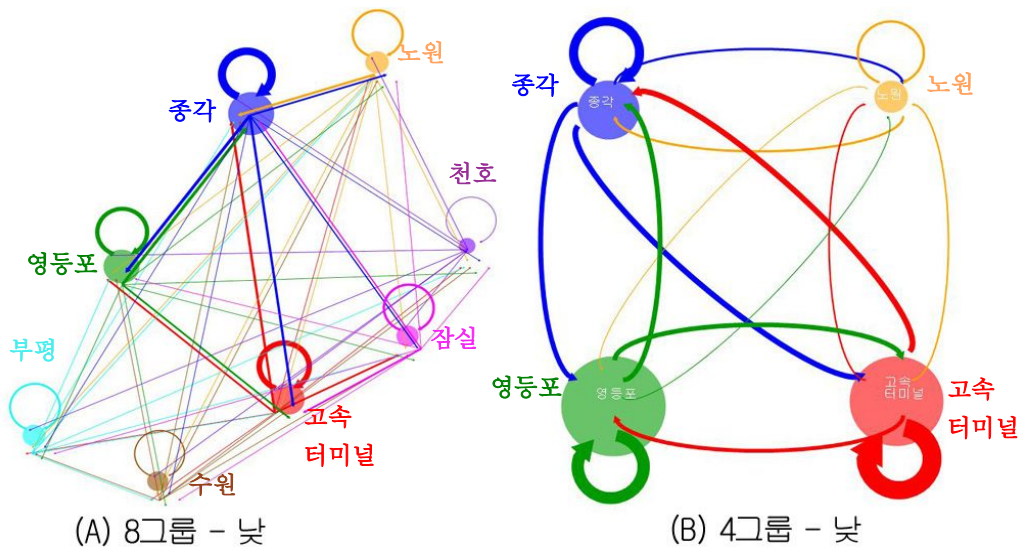
(A) 8그룹 - 아침

(B) 4그룹 - 아침

[그림 16] MST 기준 아침시간대 승객흐름도

[그림 16]은 아침시간대의 주요 역 승객흐름을 시각화 하여 나타낸 것이다. 다른 분류와 마찬가지로 MFC로 나타내었으며 (A)는 8개의 그룹, (B)는 4개의 그룹을 표현한 것이다. (A)의 8개의 그룹은 모든 그룹에서 그룹으로

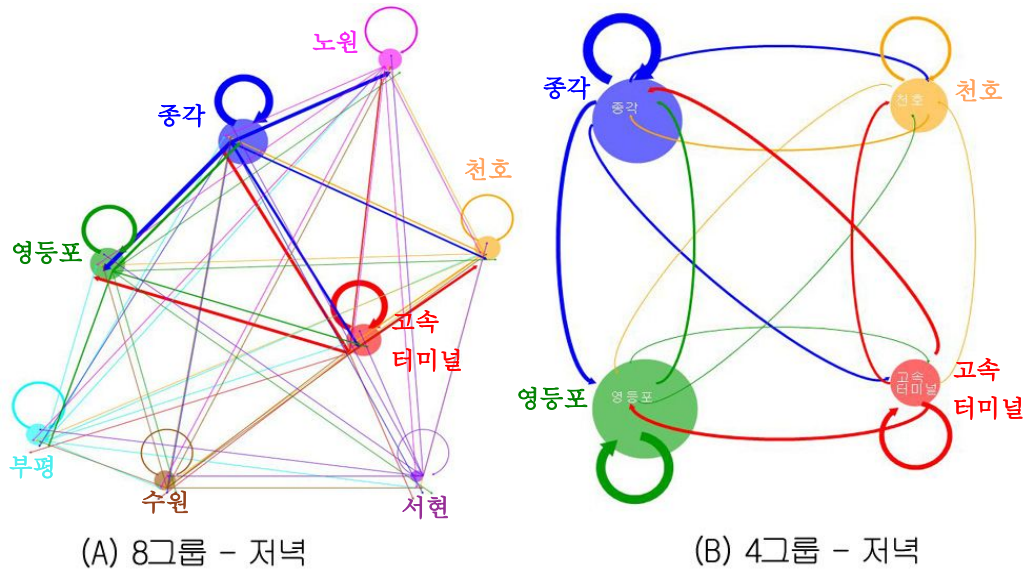
승, 하차 하는 승객수의 평균값을 구하여 평균값 이하는 선의 폭을 1로 지정하였으며, 평균값 이상은 승객 수에 비례하여 선의 굵기를 표시하였다. 또한 각 그룹을 대표하는 정점은 역시 실제 그룹의 개수로 표현하였다. 8개의 그룹의 승객 흐름을 살펴보면 가장 굵은 선은 종각역→종각역 이동그룹으로 나타났고 다음으로 신림역에서 신림역으로의 승객수가 많은 것으로 나타났다. 그 다음으로는 종각역→신림역 이동그룹, 종각역→청담역 이동그룹, 종각역→강남역 이동그룹의 그룹들은 서로 승차, 하차가 많은 것으로 보여지며, 이중 강남역을 제외한 다른 그룹 역들은 4개의 그룹으로 분류한 주요역임을 알 수 있다. 눈에 띄는 점은 신림역→종각역 이동그룹은 종각역 그룹에서 신림역 그룹으로 가는 승객 수는 적고 신림역 그룹에서 종각역그룹으로 가는 승객수가 많다는 점이다. 이는 오전 출근 시간에 신림역그룹에서 종각역그룹으로의 이동이 많기 때문으로 보여진다. (B)의 4개의 그룹으로 분류한 승객흐름을 보면 역시 종각역→종각역 이동그룹으로의 승객수가 가장 많고, 그다음 청담역→청담역 이동그룹, 신림역→신림역 이동그룹으로의 움직임이 눈에 보인다. 또한 종각역→청담역 이동그룹, 종각역→신림역 이동그룹은 서로 활발한 승, 하차가 이루어지고 있다.



[그림 17] MST 기준 낮 시간대 승객흐름도

[그림 17]은 MST 기준으로 분류한 낮 시간대의 주요 역들의 승객흐름을 시각화 한 것이다. (A)는 8개의 그룹으로 나타낸 것이고 (B)는 4개의 주요 역들의 흐름을 나타낸 것이다. 먼저 8개의 그룹들의 흐름을 살펴볼 때 종각역→종각역 이동그룹이 가장 많은 것으로 나타났으나 아침시간대와 비교해 볼 때에는 승객의 이동의 차가 다른 그룹 역들과는 크게 차이가 나지 않는 것으로 보여 진다. 그다음으로 영등포역→영등포역 이동그룹과 고속터미널역→고속터미널역 이동그룹들의 승, 하차가 많은 것으로 나타났으며 종각역→영등포역 이동그룹, 종각역→고속터미널역 이동그룹, 영등포역→고속터미널역 이동그룹, 종각역→노원역 이동그룹, 종각역→잠실역 이동그룹, 고속터미널역→잠실역 이동그룹의 움직임이 다른 그룹 역들에 비해 많은 것으로 보여 진다. 다음 (B)그림의 4개의 그룹으로 분류된 역의 승객흐름을 살펴보면, 고속터미널역→고속터미널역 이동그룹이 가장 많은 것으로 보여 지며, 영등포역→영등포역 이동그룹이 그 다음으로 활발한 것으로 나타났다. 또한

노원역 그룹을 제외한 고속터미널역→영등포역 이동그룹, 종각역→영등포역 이동그룹, 영등포역→고속터미널역 이동그룹들은 서로간의 승객들의 승, 하차가 활발한 것으로 보여 진다.



[그림 18] MST 기준 저녁시간대 승객흐름도

[그림 18]은 MST 기준으로 분류한 저녁시간대의 주요 그룹 역들의 승객 흐름도를 나타낸 것으로 (A)는 8개의 그룹, (B)는 4개의 그룹으로 나타낸 것이다. 저녁시간대에도 마찬가지로 종각역→종각역 이동그룹의 그룹이 가장 활발하며 그다음 고속터미널역→고속터미널역 이동그룹이 많은 것으로 보여 진다. 또한 종각역→영등포역 이동그룹, 종각역→고속터미널역 이동그룹, 종각역→노원역 이동그룹, 종각역→천호역 이동그룹, 영등포역→고속터미널역 이동그룹의 움직임이 다른 그룹들에 비해 많은 것으로 보여 지며 이 그룹 역들은 4개의 그룹의 주요한 그룹역이기도 하다. (B)의 승객 흐름도를 보면 종각역→종각역 이동그룹과 영등포역→영등포역 이동그룹의 승객수가

많은 것으로 보여 지며 나머지 그룹 역들은 큰 차이 없이 비슷하게 움직이고 있는 것을 볼 수 있다. 이렇게 MST기준으로 분류한 시간대별주요역들을 그룹으로 나누어 살펴보았는데, 시간대별이 아닌 이들 모두를 포함하여 전체 하루 동안의 움직임을 살펴보았다.

rank	morning	day	evening
1	강남	고속터미널	고속터미널
2	종각	부평	부평
3	수원	천호	강남
4	청담	수원	천호
5	연신내	강남	종각
6	선릉	노원	수원
7	광화문	영등포	노원
8	신림	연신내	영등포
9	양재	종각	신촌
10	마포구청	잠실	서현

rank	그룹역	역개수
1	강남	5
2	종각	52
3	수원	35
4	청담	11
5	연신내	21
6	선릉	13
7	광화문	2
8	신림	6
9	양재	3
10	마포구청	9
11	고속터미널	25
12	부평	37
13	천호	26
14	노원	37
15	영등포	44
16	잠실	12
17	신촌	4
18	서현	15

rank	그룹역	역개수
1	종각	59
2	수원	35
3	청담	11
4	연신내	25
5	선릉	16
6	고속터미널	33
7	부평	37
8	천호	26
9	노원	37
10	영등포	51
11	잠실	12
12	서현	15

[그림 19] MST기준 전체 하루 동안의 기준 역을 분류하기 위한 순서

위의 [그림 19]는 주요한 역 8개를 구하기 위한 순서이다. 중복되는 역을 제외한 모든 역 18개의 그룹 역들의 역 개수를 구한다음 그룹 역 개수가 10개 미만인 역을 제외한 12개의 역으로 그룹 역 개수를 구하였다. 그 결과 위의 [그림 19]와 같이 나왔으며, 그 12개의 그룹역중 가장 역의 개수가 많은 역 8개를 주요 역으로 선택하였고, 다시 그중 4개의 그룹으로 분류하였

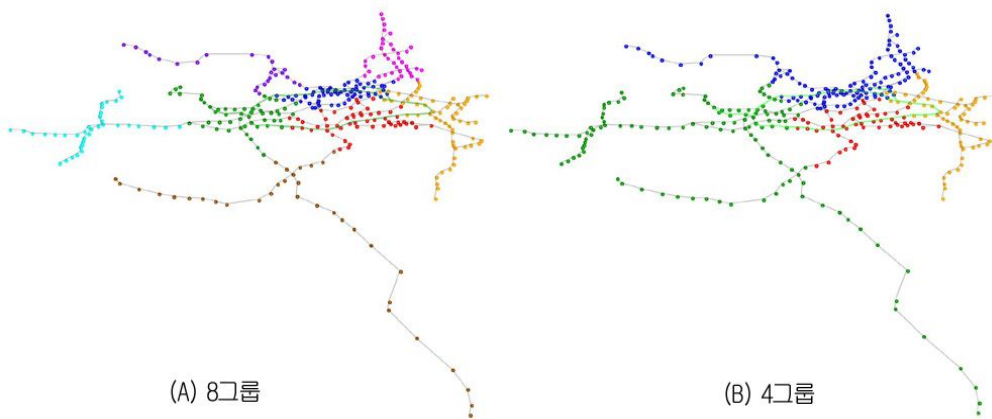
다.

아래의 [표 9]는 8개의 그룹역과 4개의 그룹역의 역 개수를 나타낸다.

[표 9] MST기준 시간대별 분류를 통한 하루 동안의 각 그룹 역 개수

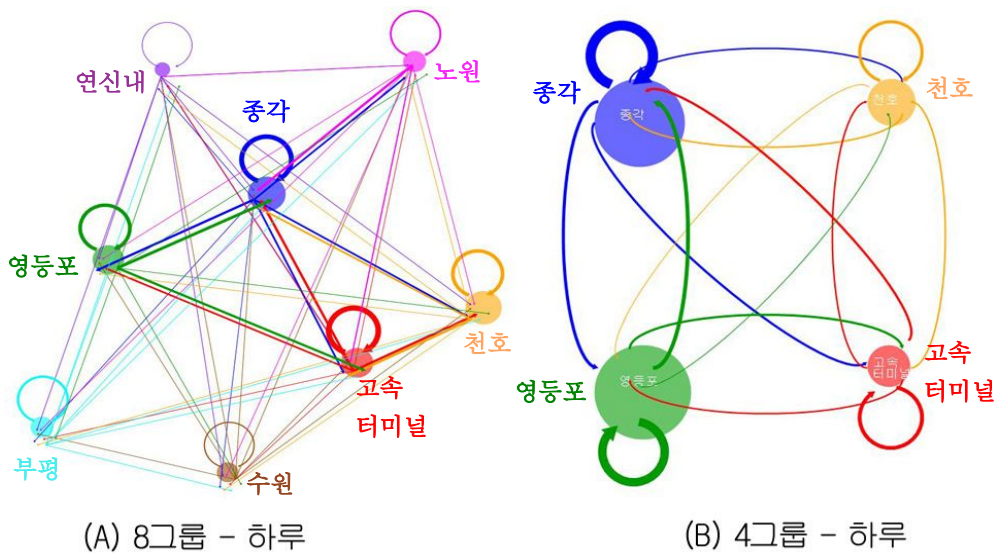
8개 그룹 역			4개 그룹 역		
	8개 그룹 역	역 개수		4개 그룹 역	역 개수
1	종각	61	1	종각	116
2	수원	35	2	고속터미널	55
3	연신내	25	3	천호	62
4	고속터미널	52	4	영등포	124
5	부평	37	X		
6	천호	59			
7	노원	37			
8	영등포	51			

8개의 그룹 역 개수를 보면 종각역과 천호역 그룹이 가장 많으나 다른 역들과의 역 개수는 큰 차이가 없는 것으로 나타났다. 다음 4개의 그룹 역 개수를 볼 때 영등포가 가장 많고 종각그룹의 역 개수가 가장 많은 것으로 나타났다. 8개의 그룹역과는 달리 다른 두 그룹과의 역 개수의 차가 큰 것으로 보여 진다.



[그림 20] MST 기준 하루 동안의 그룹 역 분포

위의 [그림 20]을 보면 주요 그룹 역들의 분포를 자세히 알 수 있다. (A)는 8개의 그룹 역을 나타내었고 (B)는 4개의 그룹 역을 나타낸 것으로써, (A)그림의 파란색은 종각그룹, 갈색은 수원그룹, 보라색은 연신내그룹, 빨간색은 고속터미널그룹, 하늘색은 부평역그룹, 노란색은 천호역그룹, 분홍색은 노원역그룹, 초록색은 영등포역그룹을 나타낸 것이다. 가장 많은 역은 종각역그룹이나 다른 그룹 역들과의 큰 차이가 없어 다른 역들과의 분포가 비슷한 것으로 보여 진다. (B)의 그림은 4개의 그룹을 나타낸 것으로써 종각역은 파란색, 고속터미널은 빨간색, 천호역은 노란색, 초록색은 영등포역그룹으로 표현한 것이다. 8개의 그룹의 분류와는 달리 초록색의 영등포역그룹과 파란색의 종각역그룹의 역 개수가 많은 것을 알 수 있다.



[그림 21] MST 기준 하루 동안의 승객 흐름도

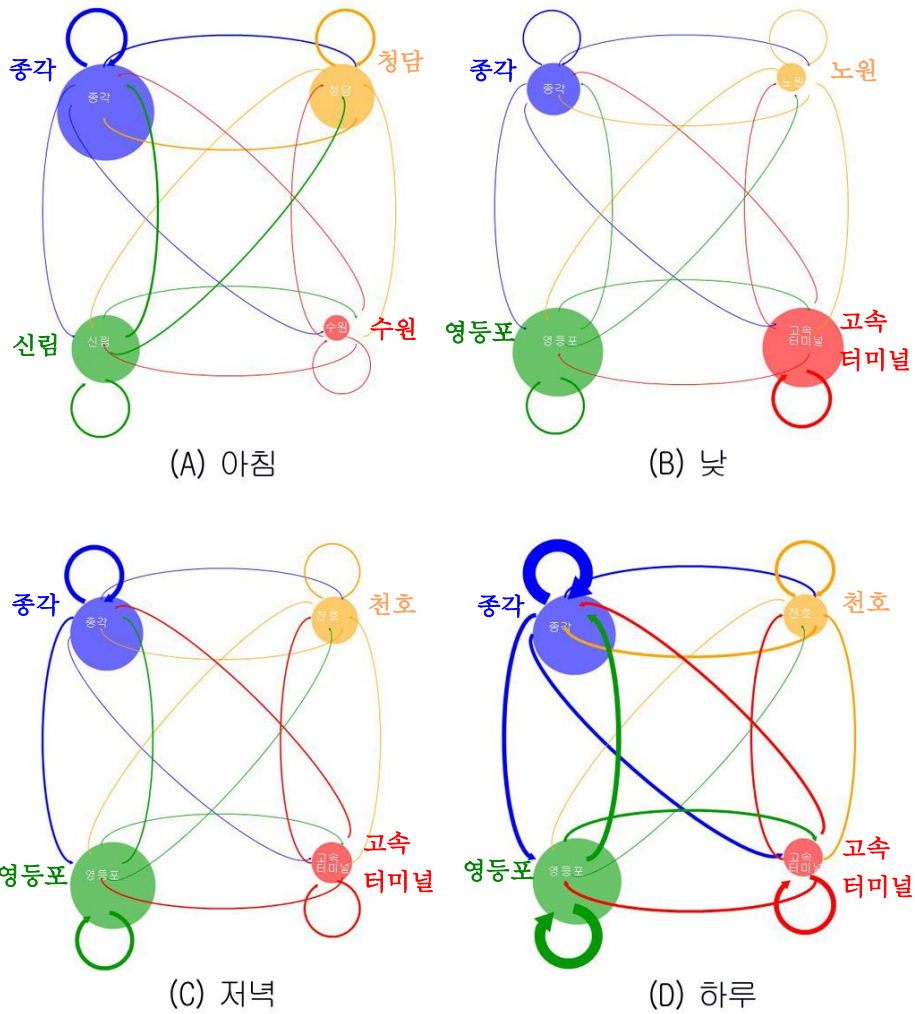
위의 [그림 21]은 MST 기준으로 시간대별 분류를 합한 전체 그룹의 승객

흐름을 MFC를 통해 나타낸 것이다. 8그룹은 기본 선 라인은 2, 전체 평균 값을 구하여 평균값 이상의 그룹은 평균값에 비례하여 두께를 그리고, 평균값 이하의 그룹은 모두 선 라인을 1로 설정했다. 4그룹의 기본 선 라인은 1이고, 전체 최소값을 구하여 그룹의 승객 수에 비례하여 두께가 그려진다. [그림 21]의 (A)는 8개의 그룹으로 나타낸 것으로 종각역→종각역 이동그룹, 고속터미널역→고속터미널역 이동그룹의 승객수가 가장 많은 것으로 보여지며 또한 (B)에서는 종각역→종각역 이동그룹이 가장 많고 그 다음으로 영등포역→영등포역 이동그룹의 그룹도 승객수가 많은 것으로 보여진다.

MST 기준으로 분류한 시간대별 분류를 합한 전체 하루 동안의 그룹을 보면 위에서 살펴보았던 저녁시간대의 그룹분류와 비슷한 것으로 보여진다. 이는 오전보다는 오후의 시간대의 그룹들이 전체시간에서 주요한 역할을 알 수 있고, 이 역들을 기준으로 승객들의 활발한 이동이 나타나는 것을 볼 수 있다.

지금까지 MST 기준으로 분류한 시간대별 승객 흐름을 알아보았는데, 이는 아침, 점심, 저녁 그리고 하루 동안의 승객 흐름도는 각자의 그룹의 특성에 맞추어서 나타낸 것이기 때문에, 이들 세 가지를 시간대별로 비교해보기 위하여 똑같은 환경에서 나타내보았다. MFC를 통하여 모든 시간대별 기본 선 폭을 1로 지정하고 기본 값을 5만 명당 1씩 증가 하도록 설정하였으며, 각 그룹을 대표하는 정점은 위에서와 마찬가지로 실제 그룹의 개수로 크기를 표현하였다.

아래의 [그림 22]는 위와 같은 설정으로 나타낸 그림이다. 아침에는 종각역 → 종각역 이동그룹, 낮 시간에는 고속터미널역→고속터미널역 이동그룹, 저녁에는 종각역→종각역 이동그룹, 영등포역→영등포역 이동그룹이 많으며, 하루 전체의 승객 흐름을 봤을 때에는 역시 종각역→종각역 이동그룹, 그리고 영등포역→영등포역 이동그룹이 가장 많다는 것을 볼 수 있다.



[그림 22] MST 기준 시간대별 승객 흐름도

마지막으로 다음 [표 10]과 [표 11]은 8개의 그룹역의 시간대별 승객수와 4개의 그룹 역 시간대별 승객수를 나타내고 있다. 위에서 시간대별로 보았던 승객흐름도와 비교하여 승객수를 볼 수 있다.

[표 10] MST기준 시간대별 분류 및 전체 8개 그룹의 승객 수

전체	출발/도착	종각	수원	연신내	고속터미널	부평	천호	노원	영등포	합계
	종각	337,440	37,516	61,066	150,655	44,828	127,317	150,064	177,969	1,086,855
	수원	40,633	102,454	2,895	34,264	8,060	6,851	7,065	45,697	247,919
	연신내	75,343	3,107	59,915	35,138	2,405	12,659	7,282	22,924	218,773
	고속터미널	156,421	30,620	28,979	374,217	24,814	155,688	62,630	144,002	977,371
	부평	50,130	8,120	2,391	28,930	173,870	7,792	5,565	70,822	347,620
	천호	136,854	6,828	11,376	172,043	7,385	261,994	50,741	53,244	700,465
	노원	171,468	7,605	7,424	77,383	5,751	54,853	144,402	32,499	501,385
	영등포	197,832	44,467	20,094	158,258	67,881	52,588	29,205	258,440	828,765
	합계	1,166,121	240,717	194,140	1,030,888	334,994	679,742	456,954	805,597	4,909,153
아침	출발/도착	강남	종각	수원	청담	연신내	선릉	신림	마포구청	합계
	강남	25,046	26,620	3,299	11,448	1,476	16,190	12,750	3,042	99,871
	종각	41,106	223,656	8,632	46,181	7,348	22,968	28,529	15,862	394,282
	수원	11,132	26,255	35,112	3,951	510	4,359	16,028	3,106	100,453
	청담	24,720	59,071	2,038	53,434	1,559	24,818	10,153	5,087	180,880
	연신내	13,694	40,553	1,057	4,657	13,616	5,740	4,122	5,526	88,965
	선릉	31,297	21,967	1,140	10,527	1,127	55,298	8,732	2,559	132,647
	신림	65,756	111,190	18,084	24,554	2,507	34,259	138,878	27,683	422,911
	마포구청	10,982	38,890	2,071	5,307	3,056	6,375	14,063	8,003	88,747
합계	223,733	548,202	71,433	160,059	31,199	170,007	233,255	70,868	1,508,756	
낮	출발/도착	고속터미널	부평	천호	수원	노원	영등포	종각	잠실	합계
	고속터미널	97,750	4,818	9,372	6,943	13,135	34,518	49,453	39,682	255,671
	부평	6,149	39,653	707	2,295	1,653	16,056	11,296	2,125	79,934
	천호	11,174	684	20,136	672	6,801	5,805	15,849	9,443	70,564
	수원	7,537	2,136	634	20,735	1,813	10,281	9,827	2,012	54,975
	노원	17,613	1,688	7,470	1,949	43,176	9,620	42,480	9,043	133,039
	영등포	38,134	15,127	5,504	10,316	8,148	77,382	60,063	13,692	228,366
	종각	47,851	9,345	13,930	8,765	34,545	54,669	131,391	25,013	325,509
	잠실	39,192	1,762	8,779	1,487	6,684	11,784	25,462	49,212	144,362
합계	265,400	75,213	66,532	53,162	115,955	220,115	345,821	150,222	1,292,420	
저녁	출발/도착	고속터미널	부평	천호	종각	수원	노원	영등포	서현	합계
	고속터미널	172,644	17,287	81,674	82,510	20,291	48,459	93,509	25,000	541,374
	부평	4,510	80,564	1,761	8,884	2,399	2,352	21,956	304	122,730
	천호	39,636	3,838	64,799	38,521	2,775	27,664	21,646	11,889	210,768
	종각	76,116	31,152	64,325	201,696	22,054	102,745	117,673	7,347	623,108
	수원	8,637	4,016	2,141	8,655	44,327	3,431	17,939	180	89,326
	노원	12,697	2,180	14,480	32,829	1,987	61,288	8,189	999	134,649
	영등포	41,837	39,262	19,680	66,864	19,302	17,798	118,402	2,987	326,132
	서현	10,759	513	10,284	4,316	222	1,954	3,288	28,554	59,890
합계	366,836	178,812	259,144	444,275	113,357	265,691	402,602	77,260	2,107,977	

[표 11] MST기준 시간대별 분류 및 전체 4개의 그룹 승객 수

전 체	출발 \ 도착	종각	고속터미널	천호	영등포	합계
	종각	949,448	256,543	196,129	330,797	1,732,917
	고속터미널	244,506	390,263	166,609	207,721	1,009,099
	천호	202,067	185,694	282,606	69,955	740,322
	영등포	354,670	229,978	69,835	772,332	1,426,815
	합계	1,750,691	1,062,478	715,179	1,380,805	4,909,153
아 침	출발 \ 도착	종각	수원	청담	신림	합계
	종각	341,798	11,053	127,167	58,047	538,065
	수원	29,242	35,112	14,111	21,988	100,453
	청담	102,072	3,770	207,679	37,804	351,325
	신림	168,766	21,498	136,570	192,079	518,913
	합계	641,878	71,433	485,527	309,918	1,508,756
낮	출발 \ 도착	고속터미널	노원	영등포	종각	합계
	고속터미널	277,653	27,514	67,441	91,298	463,906
	노원	35,338	47,050	13,585	43,951	139,924
	영등포	75,501	11,951	190,355	80,494	358,301
	종각	87,499	35,877	72,115	134,798	330,289
	합계	475,991	122,392	343,496	350,541	1,292,420
저 녁	출발 \ 도착	고속터미널	천호	종각	영등포	합계
	고속터미널	186,305	118,699	119,959	132,824	557,787
	천호	52,813	134,344	71,519	33,070	291,746
	종각	90,151	9,5812	365,356	177,513	728,832
	영등포	56,476	29,693	102,571	340,872	529,612
	합계	385,745	378,548	659,405	684,279	2,107,977

2.4 각 분류 방법의 비교 분석

다음은 본 논문에서 제시한 수도권 지하철의 모든 역을 세 가지 방법으로 분류하였을 때의 비교를 분석하여 표로 나타낸 것이다.

[표 12] 분류 방법 비교분석표

비교 방법	한강 기준 분류		K-means 알고리즘		MST 기준 분류	
	2그룹	4그룹	Method1	Method2	4그룹	8그룹
수행시간 (단위: 초)	38.72	38.924	39.938	39.576	38.768	40.164
장점	· 한강을 중심으로 수도권 지하철을 지리학적으로 분석하여 비교		· 임의의 역이 아닌 알고리즘을 이용하여 주요 역 선별		· MST를 통한 검증된 역의 선별 · 시간대별 승객 흐름을 분석하여 비교	
단점	· 임의의역 선택의 어려움		· 초기중심점의 선택에 따라 결과 값이 달라짐		· 주요 역 선택하는 방법의 어려움	
Distance 기준	정거장과 정거장 사이의 개수		실제 위치한 거리		정거장과 정거장 사이의 개수	

[표 12]를 보면 세 가지 분류 방법 모두 수행시간이 비슷한 것으로 나타났다. 그룹의 개수에 따라 약간의 시간의 차이가 난 것으로 보이며, 한강을 기준으로 분류하였을 때의 수행시간이 다른 그룹에 비해 간소한 차이로 빠른 것으로 보인다. 각 방법의 단점은 역시 중심점을 어떻게 선택하느냐에 따라 결과가 달라짐에 따라 각 분류 방법의 주요 역을 선택하는 방법들을 다양화 하여 보다 더 세분화 하는 방법을 제시하는 연구가 필요할 것이다.

V. 결 론

본 논문에서는 수도권 지하철을 교통카드의 대용량 데이터를 이용하여 수도권 지하철의 모든 역을 그룹으로 나누어 지역별 교통의 흐름과 많은 승객들의 흐름이 있는 곳의 연관성을 분석하여 수도권 대중교통의 승객 흐름을 알아보고자 하였다. 이를 위해서 2005년 6월 24일 기준 교통카드의 데이터를 이용하여 357개의 수도권 지하철역을 대상으로 분석하였다.

수도권 지하철역을 그룹으로 나누기 위해서 세 가지 방법을 선택하여 분석하였는데 그 결과는 다음과 같았다. 첫 번째, 한강을 기준으로 그룹을 나눴을 때, 강북그룹보다 강남그룹의 역 개수가 더 많았고, 같은 그룹 내에서의 이동이 활발한 것으로 보아, 수도권 지하철을 이용하는 대다수의 승객들은 출발한 역에서 목적지는 같은 지역에서 많이 움직이는 것을 알 수 있었다.

두 번째, K-means 알고리즘을 이용하여 4개의 중심점을 선택하고 그에 따른 역을 나누었을 때, 초기 중심점을 각기 다른 두 개의 방법으로 설정하여 분석한 결과, 각 4개의 그룹 역별 역 개수가 많이 차이 나며, 승객의 흐름 역시 큰 차이가 있었다. 첫 번째 방법인 한강을 기준으로 나눴을 때와는 다르게 K-means 알고리즘을 통해 나눈 그룹역이 한쪽으로 많이 몰리는 것을 볼 수 있어 그룹 역을 나누는 것은 비효과적이라고 볼 수 있었다.

세 번째, MST(Maximum flow Spanning Tree)를 이용하여 아침, 점심, 저녁대의 시간대별로 그룹 역을 나누었을 때, 역 개수는 아침, 점심, 저녁 모두 종각역의 그룹의 개수가 가장 많았고, 그중 아침 시간대의 개수가 가장 많았다. 아침시간대와 저녁시간대에는 회사가 많은 지역으로 승객의 흐름이 많은 것으로 보이며, 낮 동안에는 고속터미널역그룹이 속해있는 강남 지역의 이동이 활발한 것으로 분석된다. 시간대별 분석을 통하여 하루 전체

의 승객 이동을 분석한 결과 전체적으로는 저녁시간대의 흐름과 비슷한 것으로 나타났다. 이는 하루 중, 저녁때의 승객흐름이 가장 활발한 것으로 보인다.

본 논문에서는 수도권 지하철을 교통카드를 이용하여 승객의 흐름을 분석하였을 뿐 아니라, 기존의 연구에서 더 나아가 각각의 역 분석이 아닌 그룹으로 묶어 지역적 연계구조로 분석하였고, 이에 따른 그룹으로 나누기 위한 방법 세 가지를 제시하여 각 방법에 따른 그룹의 결과를 비교해 보았다. 또한 분석을 통해 이해하기 쉬운 결과를 보여주기 위하여 시각적으로 결과를 나타내었다. 이를 통하여 전체적으로 중심 되는 역은 종각역으로 분석되었고, 각 그룹 역은 그룹 내에서 승객의 이동이 활발함을 볼 수 있었다. 또한 오전, 오후 시간대 모두 회사가 밀집되어 있는 지역의 승객이 활발한 것을 볼 수 있었다.

위의 전체적인 분석결과로 볼 때 같은 그룹 내에서의 승객이동이 활발함에 따라 같은 지역 내의 역세권을 이용하여 지하철을 이용하는 승객들의 편리시설, 동선 등을 개선할 수 있으며, 이에 따라 수도권의 대중교통의 발전과 지역적인 토지 개발을 효과적으로 할 수 있고, 대중교통의 서비스 개선과 편리성을 도모할 수 있을 것이다. 또한 본 논문을 토대로 각 방법들에 대한 주요 역 선택을 다양화하고 세분화 한다면 보다 효과적인 정보를 얻을 수 있을 것이다. 이를 통하여 앞으로 더 나아가 수도권의 지하철 뿐만 아니라, 버스 이용과 지방의 지하철까지 연구를 확장해 나아갈 수 있다.

참 고 문 헌

- [1] 이금숙, 박종수, 김호성, 조창현, 수도권 광역도시철도 하루 시간대별 이용 빈도에 의해 구분된 역 집단과 통행자의 통행 연쇄 패턴 간 관계, 대한지리학회지, Vol.45, No.5, 2010, Pages 592~608.
- [2] Homepage of Seoul City: [Http://www.seoul.go.kr/](http://www.seoul.go.kr/)
- [3] 김은해, 대중교통체계 개편에 따른 대중교통 수요 변화 요인 분석, 서울시립대학교 석사학위논문, 2006
- [4] 김태호, 박기혁, 서울시 지하철 연계수단 실증 분석, 철도저널, Vol.13, No.2, 2010, Pages 25~29.
- [5] 박종수, 이금숙, 대용량 교통카드 트랜잭션 데이터베이스에서 통행 패턴 탐사와 통행 행태의 분석, 한국경제지리학회지, Vol.10, No.2, 2006, Pages 44~63.
- [6] 이금숙, 박종수, 서울시 대중교통 이용자의 통행패턴 분석, 한국경제지리학회지, Vol.9, No.3, 2006, Pages 379~395.
- [7] 박종수, 이금숙, 서울 대도시권 지하철망의 구조적 특성 분석, 한국경제지리학회지, Vol.11, No.3, 2008, Pages 459~475.
- [8] Chang Hyeon Joh, Daily Travel Pattern using Public Transport Mode in Seoul: An Analysis of a Multi-Dimensional Motif Search, Journal of the Korean Geographical Society, Vol.44, No.2, 2009, Pages 176~186.
- [9] 이금숙, 홍지연, 민희화, 박종수, 수도권 지하철망 상 통행흐름의 위상학적 구조와 토지이용의 관계, 한국경제지리학회지, Vol.10, No.4, 2007,

Pages 427~443.

- [10] 박종수, 이금숙, 서울 수도권 지하철 교통망에서 승객 흐름의 분석, 정보과학회논문지: 컴퓨팅의 실제 및 레터, Vol.16, No.3, 2010, Pages 316~323.
- [11] 이금숙, 민희화, 박소현, 서울시 도로교통흐름에 대한 시·공간적 분석, Vol.12, No.4, 2009, Pages 521-539.
- [12] 김호성, 박종수, 이금숙, 서울 수도권 지하철 교통망 승객 흐름의 시각화, Vol.10, No.4, 2010, Pages 397-405.
- [13] Keumsook Lee, Woo-Sung Jung, Jong Soo Park and M.Y. Choi, Statistical analysis of the Metropolitan Seoul Subway System: Network structure and passenger flows, Physica A: Statistical Mechanics and its Applications, Volume 387, Issue 24, 15 October 2008, Pages 6231-6234.
- [14] 도경구 역, "알고리즘, Foundations of Algorithms Using C++ Pseudocode", 사이텍미디어, 1999.
- [15] 용환승, 나영목, 박종수, 송현우, 이민수, 이상준, 최린 역, "데이터마이닝", 인피니티북스, 2007.
- [16] 천인국 저, "C언어로 쉽게 풀어쓴 자료구조", 생능, 2005.

ABSTRACT

Analysis of local linked structure in traffic flows of the Metropolitan Seoul

Juyoung So

Major in Computer Science Education

Graduate School of Education

Sungshin Women's University

The problem of traffic congestion in the Metropolitan Seoul area is very important for the sustainable city. To solve the problem, the Seoul Metropolitan government has been reorganizing and running the public transportation system since 2004. As a result of the reorganized public transportation system, the proportion of the bus passengers is growing, while the proportion of subway passengers is stationary. In the metropolitan area, to reduce the traffic congestion and the utilization of passenger cars, we hope that the number of subway passengers is increasing. We have focused on the analysis on the passenger flows and link structures between subway stations for the purpose of growing the subway passengers.

In this thesis, we have clustered the subway stations into some groups to analyze flows of subway passengers in the metropolitan Seoul area. In order to group all the stations we consider three different ways based on the geographical characteristics in the Metropolitan Seoul area.

Firstly, based on the Han River divided Gangnam group and Gangbuk group, it splits from the two groups to four groups again, and secondly, the K-means algorithm is used to cluster the subway stations based on random selection of the seed stations. Finally MST(Maximum flow Spanning Tree) approach is applied to analyze the flow of passengers in visual display by time interval. In addition, the passenger flows and link structures among the clustered subway stations are displayed by nodes and edges with the amount of passenger flows in different colors.

Through the analysis of the experimental results we find that the movement of passengers is very active within the same group, and also that the passengers in morning and evening are mostly crowded in the group where the companies are concentrated in. These results can be used to improve the public transportation service, and the relationship between subway station groups can be also applied to the land use for the corresponding regional development.