



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](#)

공 학 석 사 학 위 논 문

통행사슬 방법론의 신뢰성 분석

- 부산시 교통카드 기반 하차 추정률과 일치도 -

Analysis of Reliability of Trip-chain Methodology

- Estimation Ratio and Accuracy of Inferred Stop,
Using Busan Smart Card Data -

지도교수 신 용 은

2018년 2월

동의대학교대학원

도시공학과

김 수 정



공 학 석 사 학 위 논 문

통행사슬 방법론의 신뢰성 분석

- 부산시 교통카드 기반 하차 추정률과 일치도 -

Analysis of Reliability of Trip-chain Methodology
- Estimation Ratio and Accuracy of Inferred Stop,
Using Busan Smart Card Data -

지도교수 신 용 은

이 논문을 공학석사 학위논문으로 제출함

2017년 12월

동의대학교대학원

도시공학과

김 수 정

김수정의 공학석사 학위논문을 인준함

위원장

김 홍 관



위 원

김 가 야



위 원

신 용 은



2017년 12월

동의대학교 대학원

통행사슬 방법론의 신뢰성 분석

- 부산시 교통카드 기반 하차 추정률과 일치도 -

김 수 정

동의대학교대학원 도시공학과

요 약

본 연구는 하차 정보가 부재한 대중교통카드 자료로부터 하차지점 추정을 위해 적용되는 대표적 기법인 통행사슬 모형의 신뢰성 검증을 목적으로 모형의 추정 능력을 나타내는 추정률과 실제 하차 정보와의 정확성 정도를 나타내는 일치도 분석을 시도하였다. 연구 목적 달성을 위해 통행사슬 모형의 대표적인 2가지 가정을 통행사슬 Rule 1과 통행사슬 Rule 2로 구분하여 수리적 모형식으로 표현하였으며, 부산시 교통카드 자료에 모형식을 적용하여 추정된 하차 정보의 신뢰성 검증을 위한 분석을 시행하였다. Rule 1은 연속 승차 통행의 선(先) 통행 하차지점을 추정하기 위한 가정이며, Rule 2는 하루 최종 통행의 하차지점을 추정하기 위한 가정이다.

모형식은 수단통행과 정류장 수준(Stop-level)의 하차 정보를 추정할 수 있도록 설계하였으며, 정류장 수준(Stop-level) 추정을 위해 Lee et al.(2012)의 정류장 통합모형(SAM)의 제시 기준을 부산시 정류장 여건을 반영하여 수정한 후 기준에 의해 정류장 통합을 시도하였다. 모형의 알고리즘에 의해 하차 정류장 정보가 추정된 승차 통행 수를 파악하여 추정률을 분석하였고, 무료 환승 제도로 인해 실제 하차 정류장 정보가 알

려진 약 30%의 수단통행 중 추정 하차 정보가 생성된 통행을 활용하여 일치도 분석을 진행하였다. 추정률과 일치도는 정류장 간의 도보권을 고려하여 400m와 800m로 구분하여 분석하였다.

본 연구의 추정률과 일치도 분석 결과를 요약하면 아래와 같다.

첫째, 추정률을 버퍼 거리로 구분하면, 400m인 경우 Rule 1 91.0%, Rule 2 81.6% 그리고 전체 추정률은 87.7%로 파악되었고, 800m인 경우 Rule 1 95.0%, Rule 2 85.9%, 전체 추정률 91.8%로 나타났다. 버퍼 거리의 확대에 따라 추정된 하차 정류장이 증가하는 것으로 파악되었다. 전체적으로 본 연구의 통행사슬 모형 적용 시 하차 정보 추정률이 90%에 근접한 것으로 나타나 통행사슬 모형의 추정 능력이 상당히 양호한 것으로 판단되었다.

둘째, 일치도를 버퍼 거리로 구분하면, 400m인 경우 Rule 1 72.8%, Rule 2 64.7% 그리고 전체 일치도는 71.2%로 파악되었고, 800m인 경우 Rule 1 72.1%, Rule 2 63.7%, 전체 추정률은 70.5%로 나타났다. 추정률과 달리 버퍼 거리의 확대에 따라 일치도는 낮아지는 것으로 분석되었고, 본 연구의 일치도가 약 70%로 나타나 기존 실태 혹은 설문조사 자료에 의한 사실상 추정된 실제 정류장과의 비교를 통행 방법에 비해 통행사슬 모형이 보다 현실적이고 정확한 정보를 제공하여 신뢰성이 높은 것으로 판단되었다.

결론적으로 통행사슬 모형의 하차 정보 추정 능력과 추정 정보의 정확성 검증은 통해 그동안 모형 적용을 통해 알려진 연구 결과들의 신뢰성을 본 연구는 확인하였다고 할 수 있다. 향후 고도화된 통행사슬 모형의 가정 적용과 버퍼 거리뿐 아니라 통행 간 환승 시간이 포함된 다양한 알고리즘 개발이 될 경우 보다 향상된 추정률과 일치도 분석이 가능할 것으로 판단된다.

Keywords : 대중교통 시스템, 통행사슬 모형, 교통카드, 하차 추정, 정류장 통합, 빅데이터 분석

■ 목 차

제 1 장 서 론	1
1. 연구 배경	1
2. 연구 목적	3
3. 연구 범위 및 과정	4
제 2 장 이론적 고찰	6
1. 통행사슬 모형	6
1.1 개념	6
1.2 하차지점 추정 통행사슬 모형 개요	7
2. 통행사슬 모형 적용 연구	10
3. 검증 연구	12
제 3 장 분석 모형	16
1. 정류장 통합모형	16
2. 통행사슬 수리 모형 구축	19
2.1 통행사슬 Rule 1	20
2.2 통행사슬 Rule 2	23
3. 추정률 모형식	25
4. 일치도 모형	27
제 4 장 자료 수집 및 가공	29
1. 정류장 통합	29
2. 교통카드 자료 수집 및 가공	37
2.1 부산시 교통카드 개요	37
2.2 필요 속성 추출	38
2.3 자료 정리 및 가공	39

3. 분석 대상 데이터	42
 제 5 장 추정률과 일치도 분석	44
1. 통행사슬 모형 알고리즘 적용	44
2. 알고리즘 적용 결과 분석	47
2.1 추정률 결과 분석	47
2.2 일치도 결과 분석	50
 제 6 장 결 론	54
 참고문헌	58
Abstract	63
감사의 글	66

■ 표목차

<표 2-1> 추정률과 일치도에 관한 선행연구 사례	14
<표 4-1> 부산시 교통수단별 정류장 개수	30
<표 4-2> 추출 데이터 속성과 내용	38
<표 4-3> Raw Data의 자료정리 예시	39
<표 4-4> 데이터 가공 및 속성 변환	40
<표 4-5> 속성 변환 자료 예시	41
<표 4-6> 오류 자료 사유	42
<표 4-7> 분석 대상 데이터 요약	42
<표 5-1> 추정률 분석 결과	47
<표 5-2> 추정률 관련 연구	49
<표 5-3> 일치도 분석 결과	50
<표 5-4> 일치도 관련 연구	52

■ 그림목차

<그림 1-1> 연구 수행과정	5
<그림 2-1> 특정인의 하루 통행순서	6
<그림 2-2> 하차지점 추정 알고리즘	8
<그림 3-1> Integrated SAM	18
<그림 3-2> Rule 1 : 하차 정류장 추정 모형	21
<그림 3-3> Rule 2 : 하차 정류장 추정 모형	24
<그림 4-1> 통합 정류장 기준	31
<그림 4-2> 정류장 통합 및 명칭 도출 과정 예시	32
<그림 4-3> 통합 정류장 도출 과정 예시	33
<그림 4-4> 현황 및 통합 정류장 비교	35
<그림 4-5> 해운대 신도시 지역 좌표 및 통합 정류장	36
<그림 5-1> 통행사슬 모형 Rule 1 알고리즘	45
<그림 5-2> 통행사슬 모형 Rule 2 알고리즘	46
<그림 5-3> 버퍼 거리 및 Rule에 따른 추정률	48
<그림 5-4> 버퍼 거리 및 Rule에 따른 일치도	51

제 1 장 서 론

1. 연구 배경

Smart Fare Card 혹은 Automated Fare Collection(AFC) system에 의해 생성된 자료를 활용한 대중교통 노선계획 및 통행패턴 그리고 대중교통 수요 추정 등의 적용을 위한 광범위한 노력이 국내외적으로 진행되고 있다. 특히 이용자의 실제 통행 상황을 포함하고 있는 교통카드 자료로부터 생성된 기종점(O-D) 정보는 과거 표본에 의한 조사 방식과 비교한 것으로 대중교통이용자의 통행 행태에 관하여 보다 정확한 이해와 분석을 가능하게 한다. 유용한 기종점 정보의 도출을 위해서는 교통카드 자료로부터 이용자의 기점(Origin)과 종점(Destination) 정보의 추출이 가능하여야 하나, 일부 도시를 제외하고 대부분의 교통카드시스템은 승차인기점 정보만을 생성하며 하차 즉, 종점 정보는 부재한 상황이다. 이 같은 사유로 많은 연구자들이 교통카드 자료로부터 합리적이고 신뢰할 수 있는 하차 정보 추출을 위한 방법론을 찾고자 노력하여 왔다.

통행사슬 방법론(Trip-chain Method)은 하차 정보가 부재한 교통카드 자료로부터 하차지점 추정을 가능하게 하는 대표적인 모형으로 교통카드 자료에 표현된 대중교통이용자 통행에 관한 논리적인 가정을 기반으로 하차 정보를 추정하는 기법이다. 통행사슬 모형에 의해 추정된 하차 정보의 신뢰성을 전제로 대중교통계획 및 운영에 활용을 위해 추정 하차정보를 기반으로 이용자의 기종점(O-D)과 통행패턴 분석 연구가 국내외적으로 광범위하게 진행되어 왔다.

여기서 통행사슬 모형에 의해 추정된 하차 정보 신뢰성의 검증 문제는 연구자들의 주요 과제로 대두되었으나, 비교 자료와 실제 하차 정보의 부재 등으로 인해 검증 연구는 제한적으로 이루어져 왔다. 특히 시도된 검증 연구들은 추정 하차 정보와 실제 하차 정보의 일치 여부만 초점을

두어, 통행사슬 모형의 추정 규모를 나타내는 추정 능력을 검증한 연구는 전무한 실정이다. 추정된 정보의 참값 여부를 결정하는 정확성 정도가 모형의 신뢰성에 결정적인 역할을 하나, 추정된 하차 정보 규모도 모형의 신뢰성과 밀접한 관계가 있다. 즉, 통행사슬 모형에 의해 하차 정보가 도출되는 통행이 전체 통행 대비 클수록 모형의 신뢰성은 높아지며, 반대인 경우 신뢰성은 낮아진다고 할 수 있다. 통행사슬 모형의 신뢰성은 추정 정보의 정확성뿐만 아니라 추정 능력을 동시에 고려하여 검증되어야 할 것이며, 이에 따라 통행사슬 모형 적용 연구들은 추정 정보의 정확성과 더불어 추정 규모가 제고될 수 있도록 모형을 설계해야 할 것이다.

본 연구는 통행사슬 모형의 추정 능력을 추정률, 추정 정보의 정확성을 일치도로 정의하여, 관련 연구에 제시된 추정률과 일치도를 살펴보면, 각각 60.7%~83.0%, 75.0%~95.4%의 범위로 제시되어 상당한 차이를 보이고 있다. 이는 모형 적용 시 가정에 따른 추정 범위 설정¹⁾의 차이, 수단 통행과 목적통행 간의 통행량 차이 발생에 따른 추정률과 일치도의 오차 발생 그리고 교통카드 자료 간의 속성값 차이로 발생하는 것으로 판단된다. 또한, 추정률과 일치도의 계산을 위한 자료처리 방식의 차이도 사유가 되기도 한다. 특히 일치도의 경우 하차 정보가 기록된 교통카드 자료를 활용하여 존 수준(Zone-level)²⁾의 일치도를 검증한 Alsger et al.(2016)의 연구를 제외하고, 대부분의 연구가 설문조사 및 통행실태조사 등을 통해 파악된 제한된 추정 정보를 기반으로 일치도를 제시하고 있음을 지적하고자 한다.

마지막으로 대중교통이용자 통행패턴의 정밀한 분석을 위해서는 정류장 수준(Stop-level)³⁾의 하차지점 도출이 필수적이다. 즉, 존 수준

-
- 1) 추정 범위는 단일지점인 정류장부터 도보권 및 행정동 내에 위치하고 있는 여러 개의 정류장들을 통합한 하나의 교통권역, 행정동 등 다양하게 설정할 수 있음
 - 2) 존 수준(Zone-level)은 추정 범위에 의해 설정된 구역으로 단일 정류장을 의미함
 - 3) 본 연구의 정류장 수준(Stop-level)은 거리, 정류장 명칭 등 단일 정류장으로 통합된 정류장을 의미함

(Zone-level)에서 추정된 하차 정보는 대중교통이용자의 개략적인 기종점 및 통행패턴 정보만을 제공하기 때문에 노선계획 및 운행계획과 같이 하차 정류장 정보가 요구되는 분야에 적용이 제한적으로 이루어질 수 밖에 없다. 반면 통행사슬 모형은 정류장 수준(Stop-level)의 하차 정보 도출이 가능하나, 실제 하차 정보 자료수집의 한계⁴⁾로 인하여 정류장 수준(Stop-level)의 추정률과 일치도 검증에 관한 연구는 부재한 상황이다. 따라서 통행사슬 모형을 정류장 수준(Stop-level)의 하차지점 추정이 이루어지도록 설계하여 추정된 하차 정류장 정보를 기반으로 추정률과 일치도의 검증이 이루어져야 할 것이다.

2. 연구 목적

본 연구는 하차 정보가 부재한 대중교통카드 자료로부터 하차지점 추정을 위해 적용되는 대표적 기법인 통행사슬 모형의 추정 능력을 나타내는 추정률과 실제 지점과의 일치 정도를 나타내는 일치도 분석을 통해 모형의 신뢰성을 검증하고자 한다. 구체적으로 기술하면 아래와 같다.

첫째, 정류장 수준(Stop-level)의 통행사슬 기반 모형을 구축하고, 추정률과 일치도 분석을 위한 산정모형식을 도출한다. 통행사슬 모형은 선(先) 통행의 하차 정류장 추정을 위한 가정과 하루 최종 통행의 하차 정류장 추정을 위한 가정을 수리적인 식으로 표현되도록 설계한다.

둘째, 도출된 산정모형식을 이용하여 교통카드 자료의 추정률과 일치도를 산출함으로써 모형의 신뢰성을 검증한다. 추정률 일치도의 신뢰성 검증은 부산시 교통카드시스템 생성 자료를 활용한다.

마지막으로 정류장 수준의 하차정보 도출을 위해 Lee et al.(2012)의 SAM(Stop Aggregation Model)을 준용하여 부산시 대중교통 정류장을 통합하고, 수집된 부산시 교통카드 자료를 정리 및 가공하여 구축된 모

4) 거리비례제를 시행하는 서울시와 달리 부산시의 경우 무료 환승 혜택으로 인해 환승을 위한 교통카드 태그로만 하차지점을 알 수 있음

형식을 적용한다.

3. 연구 범위 및 과정

하차 정류장 추정을 위한 통행사슬 기반 모형 제시와 더불어 모형의 신뢰성 검증을 위해 <그림 1-1>과 같은 과정을 통한 연구를 수행하였으며, 모형의 실증적인 검증을 위하여 2016년 기준의 부산시 대중교통(도시철도, 시내버스, 마을버스)카드의 승하차 자료를 활용하였다.

제 1 장은 연구의 배경과 목적을 기술하였다.

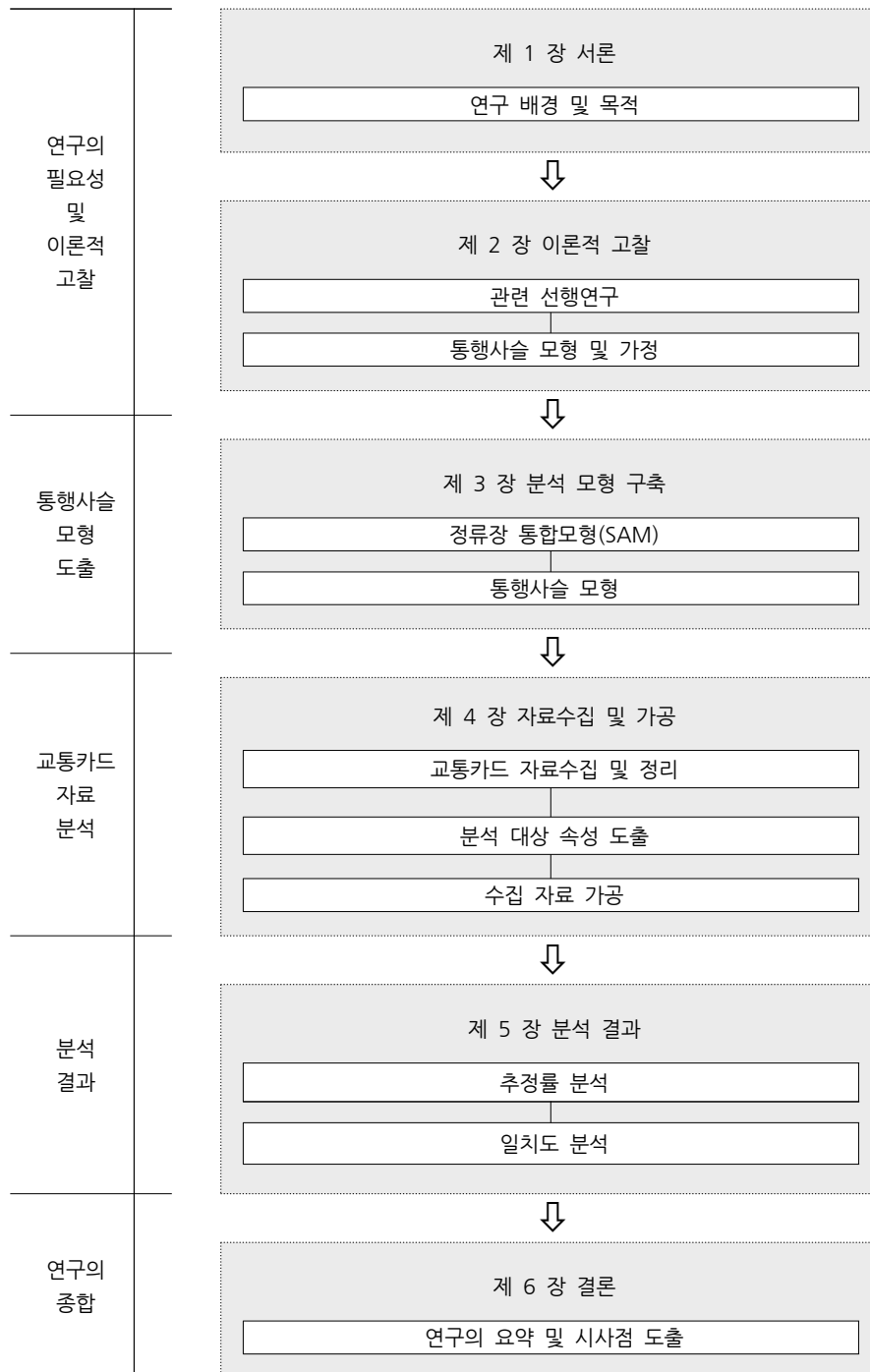
제 2 장은 통행사슬 모형에 관한 전반적인 이론적 고찰을 거쳐 통행사슬 모형 적용 및 신뢰성 검증을 시도한 선행연구들을 살펴보았다.

제 3 장에서 본 연구에 적용될 모형 구축을 시도하였다. 우선 정류장 수준(Stop-level)의 하차지점 추정을 위해 Lee et al.(2012)의 SAM을 준용한 정류장 통합모형을 기술하고, 통행사슬 모형의 대표적 2가지 가정과 본 연구의 목적인 추정 규모와 일치도 분석을 위한 산정모형식을 도출하였다.

제 4 장은 도출된 산정모형식 적용을 위해 수집된 부산시 교통카드 자료로부터 분석에 필요한 속성을 도출하기 위한 자료 가공 과정을 기술하였으며, 자료 가공 과정에서 파악된 분석이 불가능한 오류자료에 관한 내용도 기술하였다.

제 5 장에서는 제 3 장에서 구축한 정류장 통합모형과 통행사슬 기반 모형을 적용한 추정 하차 정류장 정보를 활용하여 추정률과 실제 하차 정류장과의 비교를 통해 파악된 일치도 결과를 분석하였다.

제 6 장은 본 연구의 결론부로서 분석 결과를 요약 및 정리한 후, 시사점과 향후 연구 방향에 대하여 기술하였다.



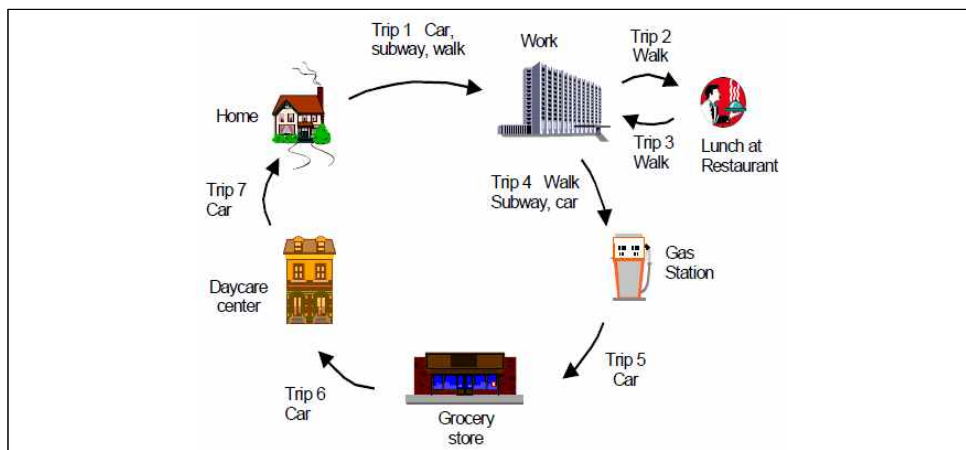
〈그림 1-1〉 연구 수행과정

제 2 장 이론적 고찰

1. 통행사슬 모형

1.1 개념

통행사슬(Trip-chain 혹은 Trip-chaining) 모형은 미국의 가구통행 실태 조사인 NHTS(National Household Travel Survey)의 자료에서 획득한 개인의 연속된 통행 정보로부터 통행을 사슬화(chaining)하여 목적통행의 기점과 종점을 파악하는 방법론이다. 통행사슬은 수단통행으로 구분된 각 통행을 논리적으로 사슬화함으로써 주요 기종점(anchor) 간의 통행을 파악하는 기법 혹은 과정을 의미한다. 비록 통행사슬(Trip-chain)의 공식적 정의는 부재하나, 일반적으로 사슬화된 통행(chained-trip)은 연속적인 통행으로 구성되며, 중간 기착지도 포함된 통행으로 이해할 수 있다. McGuckin(McGuckin N. et al., 2004)은 아래 <그림 2-1>의 연속된 통행의 예시로부터 'Chain' 및 'Tour'의 개념을 설명한다.



자료 : McGuckin, N. and Nakamoto Y., "Trips, Chains, and Tours-Using an Operational Definition," The NHTS Conference, (2004), p.5.

<그림 2-1> 특정인의 하루 통행순서

<그림 2-1>에서 제시된 개인의 하루 통행은 7개의 수단통행으로 구성

되어 있다. 7개의 통행은 집과 직장 간 두 개의 목적지(anchor) 사이에서 발생하는 첫 번째 통근 통행이며, 비 통근 통행인 점심 식사를 위한 2개의 도보 통행, 마지막으로 퇴근을 위한 1개의 귀가 통행으로 구성된다. 마지막 귀가 통행의 경우는 4개의 수단통행이 사슬화되어 1개의 목적통행으로 간주된다. 본 예시는 7개의 수단통행이 사슬화될 경우 3개의 목적통행이 됨을 보여준다.

McGuckin(2004)은 이렇게 사슬화되어 형성된 통행을 'Tour', 통행의 최종 목적지를 'Anchor', 주유소, 식료품 가게, تاک소 등은 'Intervening Stop'으로 정의한다. 'Tour'는 우리나라의 목적통행과 동일한 의미이며, 마지막으로 'Chain'은 주유소, 식료품 가게, تاک소의 연속 통행인 4개의 통행이 1개의 목적통행으로 사슬화하는 것을 말한다. 연속된 통행의 사슬화 기준은 지점 간 거리 혹은 시간이며, 일정거리 혹은 일정 시간을 초과하는 경우 사슬화하지 않는다.

1.2 하차지점 추정 통행사슬 모형 개요

Barry et al.(2002)는 하차 정보가 부재한 미국 뉴욕 'Metro Card' 자료로부터 최초로 통행사슬 개념을 적용한 하차지점 추정을 통행 뉴욕 지하철 이용자의 역간 O-D 표를 도출하였다. Barry(2002)의 연구 이후 통행사슬 모형을 통해 교통카드 자료로부터 대중교통이용자의 O-D를 포함한 다양한 통행패턴을 도출하기 위한 연구가 광범위하게 진행되어 있다.

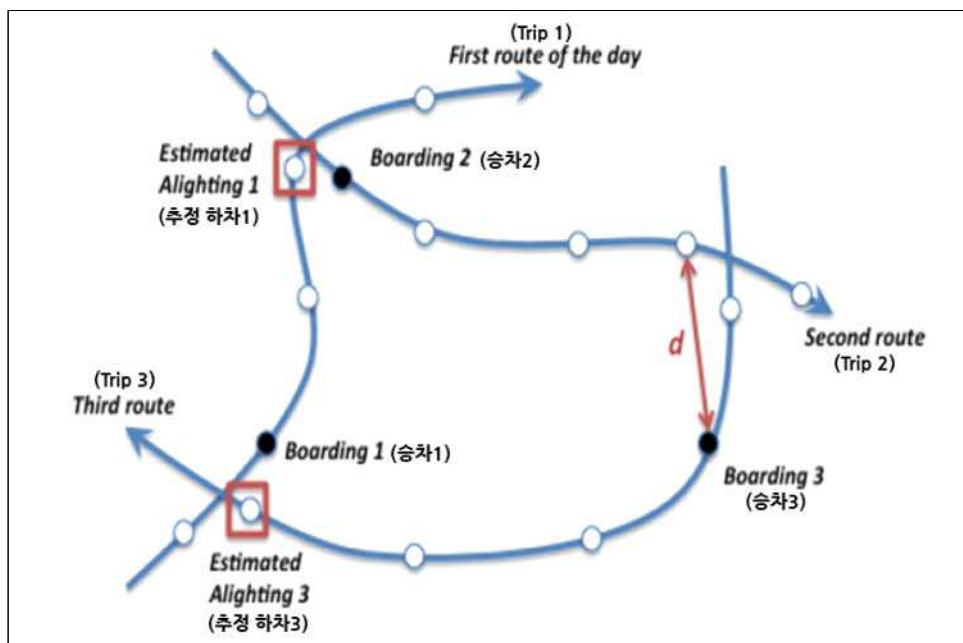
대중교통카드 분야에 적용되는 통행사슬 모형의 알고리즘들은 기본적으로 사슬화를 위한 가정들 하에 이용자의 하루 통행 순서에 따라 하차지점을 추정하도록 구성되어 있다. Barry(2002)가 제시한 두 가지 가정은 다음과 같다. 첫째, '모든 통행자는 이전 통행의 하차지점으로 돌아온다', 둘째, '하루 마지막 통행은 그날 첫 번째 통행의 승차지점으로 돌아온다'. 이후 Barry et al.(2009)는 위 두 가지 가정을 다음과 같이 확대하였다.

첫째, '대부분의 통행은 이전 통행의 하차지점 혹은 인근지점에서 다음

통행을 시작한다.'

둘째, '대부분의 통행은 하루 첫 번째 통행의 승차지점 혹은 인근지점에서 그날의 마지막 통행을 끝낸다.'

<그림 2-2>는 어느 개인의 도식화된 하루 3개의 연속 통행으로부터 언급된 통행사슬 가정에 따라 하차지점 추정 과정을 제시하고 있다.



자료 : He, I., Nassir N., Trépanier, M., Hickman, M., "Validating and Calibrating a Destination Estimation Algorithm for Public Transport Smart Card Fare Collection Systems", No. CIRRELT-2015-52, (2015), p.3.

<그림 2-2> 하차지점 추정 알고리즘

각 통행의 하차지점 추정은 다음과 같이 이루어진다. 첫 번째 통행의 승차지점이 Boarding 1이며, 두 번째 통행의 승차지점은 Boarding 2로 표기되어 있다. 통행사슬 가정에 의하면 첫 번째 통행의 하차지점은 두 번째 통행의 승차지점(Boarding 2)과 가장 가까운 첫 번째 통행노선의 지점(Estimated Alighting 1)이 된다. 마찬가지로 두 번째 통행의 하차지점은 세 번째 통행의 승차지점(Boarding 3)과 가장 가까운 두 번째 통행

노선의 지점이 된다. 수단통행과 목적통행의 구분을 위해서 통행 간 일정 시간 혹은 정류장 간 일정 거리가 기준으로 적용될 수 있다. 예를 들면, 세 번째 수단통행의 승차지점과 추정된 하차지점 간의 거리 (d)가 일정 기준을 초과하면 두 번째 통행의 하차지점은 추정 불가능한 것으로 판단한다. 목적통행의 과정에서는 두 번째 통행과 세 번째 통행은 사슬화되지 않은 별개의 통행으로 판단하며, 만약 일정 거리 이내라면 2개의 통행은 사슬화된 1개 통행으로 결정된다.

마지막으로 하루 최종 통행인 세 번째 통행의 경우 첫 번째 통행의 승차지점(Boarding 1)과 가장 가까운 세 번째 통행 노선의 지점(Estimated Alighting 3)이 되며, 정류장 간 거리가 일정 거리를 초과하는 경우 추정이 불가하다.

Barry(2002)의 두 가지 가정을 수정 혹은 확대하고 각 도시의 여건을 고려하는 등 여러 연구가 통행사슬 모형을 고도화하여 왔으나, 다음 두 가지 가정은 여전히 공통적으로 적용되고 있다.

첫째, 이전 통행의 하차지점이 다음 통행의 승차지점과 연관이 있다는 연속된 통행에 관한 가정과,

둘째, 하루 마지막 통행의 하차지점 추정에 적용되는 마지막 통행의 하차지점은 첫 번째 통행의 승차지점과 연관이 있다는 가정이다.

연관성을 판정하는 기준은 승하차지점 간 거리 혹은 통행 간 통행 시간 차이이며, 이에 따라 하차지점 추정의 가능 여부가 결정된다. 또한, 통행 노선들의 동일성 여부는 추정 가능성 및 용이성과 밀접한 관련이 있다.

사정의 다양성을 살펴보면, 마지막 통행의 하차지점이 다음 날의 첫 번째 승차지점이 된다는 가정도 있으며, 다양한 환승 거리 및 환승 시간에 대한 가정 하에 진행된 연구도 다수이다. 그러나 본 연구는 위에 제시된 대표적 2가지 가정만을 적용하고자 한다.

2. 통행사슬 모형 적용 연구

Barry et al.(2002) 이후 가정을 추가하거나 기법을 고도화한 연구가 세계 여러 도시의 교통카드 자료를 활용하여 진행되어 왔다. 몇몇 대표적인 연구를 살펴보면 아래와 같다.

Trepanier et al.(2006, 2007)는 사슬화를 위한 최소 거리 가정의 단일 통행 하차지점 추정을 위한 전날 통행의 하차지점이 다음날 단일 통행의 하차지점과 동일하다는 가정을 추가한 통행사슬 모형을 캐나다 퀘벡 Gatineau 시의 스마트카드 자료에 적용하였다. Munizaga and Palma(2012)는 Trepanier et al.(2007) 연구와 유사 모형을 적용하고, 추정값의 일치도 제고를 위해 각 통행의 방향별 하차지점을 감안하여 하차지점 추정을 시도하였다.

효율적이고 신뢰성 높은 목적통행의 하차지점 추정을 위해 연구에 따라 차별적인 환승 시간과 거리가 가정되었다. 환승 거리를 가정한 연구로는 Cui(2006), Wang(2011), Nassir et al.(2011) 및 Munizaga와 Palma(2012)를 들 수 있으며, 환승 시간을 가정한 연구는 Bagchi와 White(2004), Hofmann and O'Mahony(2005), Nassir et al.(2011), Devillaine et al.(2012), Kieu et al.(2013), 및 Ma et al.(2013)가 해당된다. 예를 들면, Devillaine(2012)는 Santiago와 Gatineau 시의 대중교통이용자의 통행패턴과 목적 파악을 위한 기초적 정보 추출을 위해 통행사슬 모형을 적용하여 하차지점을 추정하면서 동일 노선인 경우 30분 환승 시간을 가정하였고, Chu and Chapleau(2008)은 5분을 환승 시간으로 가정하였으며, Nassir et al.(2011)는 최소 30분과 최대 90분으로 설정하여 분석을 진행하였다.

추정의 용이성과 일치도 제고를 위해 교통카드 자료와 더불어 보완적인 교통 정보 자료를 활용한 연구도 방대하다. 보완적 정보자료로는 버스정보시스템(Bus Information System)에서 도출된 AVL(Automatic Vehicle Location) 시스템 정보, 차량 GPS 정보, 혹은 APC(Automatic

Passenger Counter) 정보 등이 여기에 해당된다. 관련 연구로는 Cui et al.(2006), Zhoa et al.(2007), Nassir et al.(2011), Wang et al.(2011) Munizaga et al.(2012, 2014) 등을 들 수 있으며, 대부분 보다 정밀한 대중교통이용자의 O-D 도출 혹은 통행패턴 파악을 위해 언급된 시스템의 생성 자료와 교통카드 자료의 통합 방법론과 이를 활용한 추정 방법론을 제시하였다. 예를 들면, Wang et al.(2011)은 런던의 iBus Data(AVL and radio system)와 교통카드인 Oyster 카드 정보의 통합 방법론을 적용하여 추정 O-D 표를 도출하고, 실제 통행 조사 자료에서 도출된 O-D 표를 비교·분석하여 제시된 방법론을 검증하였다.

광범위한 국외 연구가 진행되어온 반면, 국내의 경우는 관련 연구가 제한적이다. 김순관 등(2007)은 서울시 교통카드 자료에 기본적인 통행사슬 모형의 가정을 적용하여 하차정보가 결측된 데이터의 3가지 개인 통행의 보정 방안을 제시하였다. 첫째, 연속된 통행의 첫 번째 통행 하차 보정 방안, 둘째, 연속된 날짜의 반복되는 개인의 통행 정보를 활용하는 방안, 마지막으로 통행의 회귀적 특성을 활용한 보정 방안을 제시하였다. 그 외에 노선별 보정 방안 및 총량적 보정 방안 등을 추가하여 상세한 하차지점을 도출하고자 하였다. 조아 등(2015)은 청주시 교통카드 자료에 LDA(Latent Dirichlet Allocation) 기법을 적용하여 시내버스 이용자의 통행패턴을 파악하였다. 패턴 파악의 핵심 사안은 하차 추정으로서 통행사슬 모형의 기본적 가정을 적용하였다. 마지막으로 신강원(2016)은 부산시 선불 교통카드 자료를 이용하여 통행사슬 모형의 두 가지 가정과 환승거리 500m 가정을 적용하여 대중교통이용자의 존별 통행 종점 추정을 시도하였다. 하루 2 통행 이상 이용자 중 93.7%가 통행 종점 추정이 가능한 것으로 제시하였고, 실제 종점과의 일치도는 78.5%로 분석하였으며, 일치도는 환승 거리 가정에 따라 변화하는 것으로 제시하고 있다.

3. 검증 연구

앞서 살펴보았듯이 교통카드 자료를 기반으로 통행사슬 모형을 적용하여 하차지점 추정을 시도한 연구는 다양한 측면에 방대하게 진행되어 왔으나, 본 연구의 목적인 통행사슬 모형 적용에 따른 추정 규모와 추정 값의 신뢰성(즉, 실제 하차지점과의 일치성 여부)에 관한 연구는 매우 제한적으로 이루어져왔다. 그 사유는 다양하나 대부분 실제 하차 정보를 제공할 수 있는 자료 혹은 정보의 한계가 가장 큰 이유이다. 특히 통행사슬 모형의 추정 규모 파악을 목적으로 연구된 경우는 전무하며, 일부 연구에서 추정률만 언급돼있을 뿐이다. 반면 실제 하차지점과의 일치 여부를 나타내는 정확성의 검증을 주제로 진행한 연구는 제한적으로 이루어졌다.

검증 연구의 대부분은 실제 하차지점 파악을 위해 현장 조사 혹은 설문조사를 활용하는 경우가 대부분이다. 그러나 현장 혹은 설문조사는 전수 조사가 불가능하기 때문에 불가피하게 일부의 결과와 비교하여 검증할 수밖에 없는 한계가 있다. Barry et al.(2009)는 통행사슬 모형 적용으로 도출된 추정 정류장을 지하철 출입구의 승하차 수와 버스 정류장 및 지하철 출입구에 집계원을 배치하여 직접 수집한 승하차 정보를 비교하여 검증을 시도하였다. 이 경우 정류장별 총체적인 승차와 하차의 개략적인 검증은 가능하나 개인 통행별 하차지점의 일치 여부 파악은 불가능하다고 할 수 있다.

통행사슬 모형을 적용하지 않았으나, 설문조사 정보와 교통카드 자료 간의 일치도 검증을 위해서 Chow(2014)는 일반적인 설문 방식이 아닌 온라인 이용자 설문조사 방법을 제시하였고, 특히 Prompted Recall 조사 방법 적용 시 응답 자료와 교통카드 자료 간 일치 비율이 거의 67.3%에 이른다고 분석하였다. Munizaga et al.(2014)는 통행사슬 방법론의 신뢰성 검증을 목적으로 기법 상의 가정들 및 추정 과정들에 대한 전반적인 검증 방법을 제시하였으며, 특히 Santiago metro 역간 O-D 조사 자료와

통행사슬 모형에 의해 추정된 하차지점으로부터 도출된 O-D 표 간의 비교를 통해 약 84.2%가 일치도를 제시하였으나 제시된 수치가 지하철 수단에 한정되었다는 단점이 있다.

실제 하차 정보가 주어진 교통카드 자료를 획득하여 통행사슬 모형의 신뢰성을 검증한 대표적인 연구는 Alsger et al.(2016)이다. Alsger(2016)는 승차와 하차지점 정보가 모두 주어진 호주 SEQ 대중교통 공사 TransLink의 교통카드 자료를 이용하여 추정 값의 일치도 검증을 시도한 연구로서 본 연구의 목적과 가장 근접한 내용을 포함하고 있다. 즉, 하차 정보가 주어진 교통카드 자료와 본 연구와 유사한 가정 하에 통행사슬 모형을 적용하여 하차지점을 추정하고 실제 하차지점과 비교하여 일치도를 평가하였다. Alsger(2016)의 연구는 버스, 도시철도, 페리를 구성하는 다수단 대중교통 통행자의 정보가 포함된 자료를 활용하고, 환승시간 혹은 거리의 가정에 따라 도출된 하차 추정 값의 일치도에 대한 민감도 분석을 시도하였다는 장점이 있다.

신뢰성 검증은 각 통행의 실제 O-D와 추정 O-D를 비교하여 동일한 기종점 정류장/존을 가진 통행의 일치(매칭) 비율(총 통행 대비 실제와 추정이 동일한 기종점 통행을 나눈 값)의 분석을 통해 이루어졌다. 1,515개로 구성된 존 수준(Zone-level)의 경우 일치 비율이 72.6%로 분석되었으며, 평균 오차 (실제 존(Zone)과 추정 존 수준(Zone-level) 차이)는 약 530m로 계산되었다.

본 연구와 달리 Alsger(2016)의 연구는 목적통행과 존 수준(Zone-level) 추정에 초점을 두어 환승거리 800m 혹은 환승시간 60분 이내를 가정하여 연구를 진행하였다. 정류장 수준(Stop-level)의 일치 비율도 파악하였으나, 환승거리 800m 및 환승시간 30분, 60분, 90분을 각각 가정한 결과 모두 60% 이하의 일치 비율로 나타났다.

아래의 <표 2-1>에 관련 연구들이 요약되어 있으며, Jung and Sohn(2017)의 연구를 참고하여 연구 대상 수단, 데이터, 추정률, 일치도, 검증 데이터 등도 요약하였다.

〈표 2-1〉 추정률과 일치도에 관한 선행연구 사례

저자	교통수단	포함된 데이터	추정률	신뢰도	검증 데이터
Barry et al.(2002)	지하철	-	-	90%	통행실태조사
Trepanier et al.(2006)	버스	오류 1통행 2-km	66%	-	-
Zhao et al.(2007)	지하철-지하철 지하철-버스	오류	71.2%	-	-
Chapleau et al.(2007)		Rule1만 적용	81% (300m) 92% (1km)	-	-
Lianfu et al.(2007)	버스	-	-	-	-
Reddy et al.(2008)	버스, 지하철	다음날 통행 오류 1통행	-	90%	통행 설문조사
Barry et al.(2009)	버스, 지하철, 트램, 페리	-	-	-	-
Munizaga et al.(2010)	버스, 지하철	오류 1통행	82%	-	-
Li et al.(2011)	버스	1통행	70% (전일) 80% (첨두)	-	-
Nassir et al.(2011)	버스	-	60.74% (승하차 모두)	95.4%	-
Wang et al.(2011)	버스	-	66% Northbound 65% Southbound	-	-
Munizaga et al.(2012)	버스, 도시철도	오류 1통행 같은 위치	80.77% (2009년 3월) 83.01% (210년 6월)	-	-

〈표 계속〉

저자	교통수단	포함된 데이터	추정률	신뢰도	검증 데이터
Munizaga et al.(2014)	버스-도시철도 도시철도-버스	오류 (1km, 30분)	-	84.2%	53 지원자 (주로 학생)
Zhang et al.(2015)	버스	-	-	75%	무료 배포 교통카드 (102명)
Alsger et al.(2015)	버스, 철도, 페리	-	-	-	-
He et al.(2015)	Brisbane 교통수단	-	-	65.8% (0m)	실제 하차 데이터
				79% (400m)	
				85% (1000m)	
Nunes et al.(2016)	버스	오류	62.4%	-	실제 하차 데이터
Alsger et al.(2016)	버스, 철도, 페리	-	-	72.6% (800m)	실제 하차 데이터

제 3 장 분석 모형

본 장에서는 교통카드 자료상의 각 승차 통행(수단통행)에 대한 정류장 수준(Stop-level)의 하차지점 추정에 초점을 두고 통행사슬 모형의 하차 정보 추정 능력과 추정된 하차 정보의 정확성 분석을 통한 모형의 신뢰성 검증을 위해 수리적인 분석 모형을 구축한다. 모형 구축 과정은 아래와 같이 요약할 수 있다.

첫째, 정류장 수준(Stop-level)의 추정을 위한 정류장 통합을 위해 Lee et al.(2012)가 제시한 SAM(Stop Aggregation Model)을 준용한다.

둘째, 통행사슬 모형의 대표적인 2가지 가정의 수리 모형식을 도출한다.

셋째, 추정률과 일치도 분석을 위한 모형식을 도출한다.

1. 정류장 통합모형

단일 정류장(역)이 이용자의 모든 방향별 승하차를 처리하는 도시철도 수단과 달리 노선 유연성이 장점인 시내버스와 마을버스 수단은 방향별 정류장의 설치뿐만 아니라 다수의 노선이 단일 정류장을 공유하거나 동일 방향임에도 인접한 곳에 복수의 정류장 설치가 요구되기도 한다. 대중교통계획 관점에서 통행 기종점은 개별 정류장 위치가 아닌 정류장 인근 지역이 되기 때문에 방향별 혹은 다수 정류장의 개별 위치는 무의미하며, 동일한 접근 권역과 연계된 것으로 판단되는 복수의 정류장들은 1개 정류장으로 간주되어야 한다. 따라서 본 연구가 시도하는 정류장 수준(Stop-level)의 하차지점 추정을 위해서는 공간적으로 분리된 시내버스 및 마을버스 정류장들을 적정 기준에 따라 새로 통합된 정류장으로 표현하여야 한다.

Lee et al.(2012)는 언급한 상황의 해결을 위해 정류장 통합모형(Stop Aggregation Model - SAM)을 제시하였다. SAM의 기본 가정은 대중교

통 통행의 기종점은 정류장 위치가 아닌 정류장 인근에 위치한 통행유발원(즉, 주택, 직장, 마트 등)이며, 통행유발원은 여러 대중교통수단의 정류장들에 의해 접근이 가능할 수 있다는 것이다. SAM의 궁극적인 목적은 개별 정류장에 비해 대중교통이용자 기종점인 통행유발원의 특성을 보다 근접하게 정의하는 일반화된 정류장을 새로이 만드는 것으로, 예를 들면 인근 지역의 Landmark, 대형 교통유발원인 백화점 및 주요 교차로로 정의하거나, 혹은 여러 특성을 복합적으로 나타낼 수 있는 통합된 새로운 정류장으로 정의하는 것으로 지적하였다.

Lee(2012)는 정류장 통합의 일반적 기준으로 다음 3가지를 제시하였다.

- 정류장 간 거리 혹은 인접 정도(Distance-Based SAM : DBSAM)
- 정류장 명칭의 유사성(Text-Based SAM : TBSAM)
- 토지이용의 유사성(Catchment-Based SAM : CBSAM)

각 기준을 요약하면 아래와 같다.

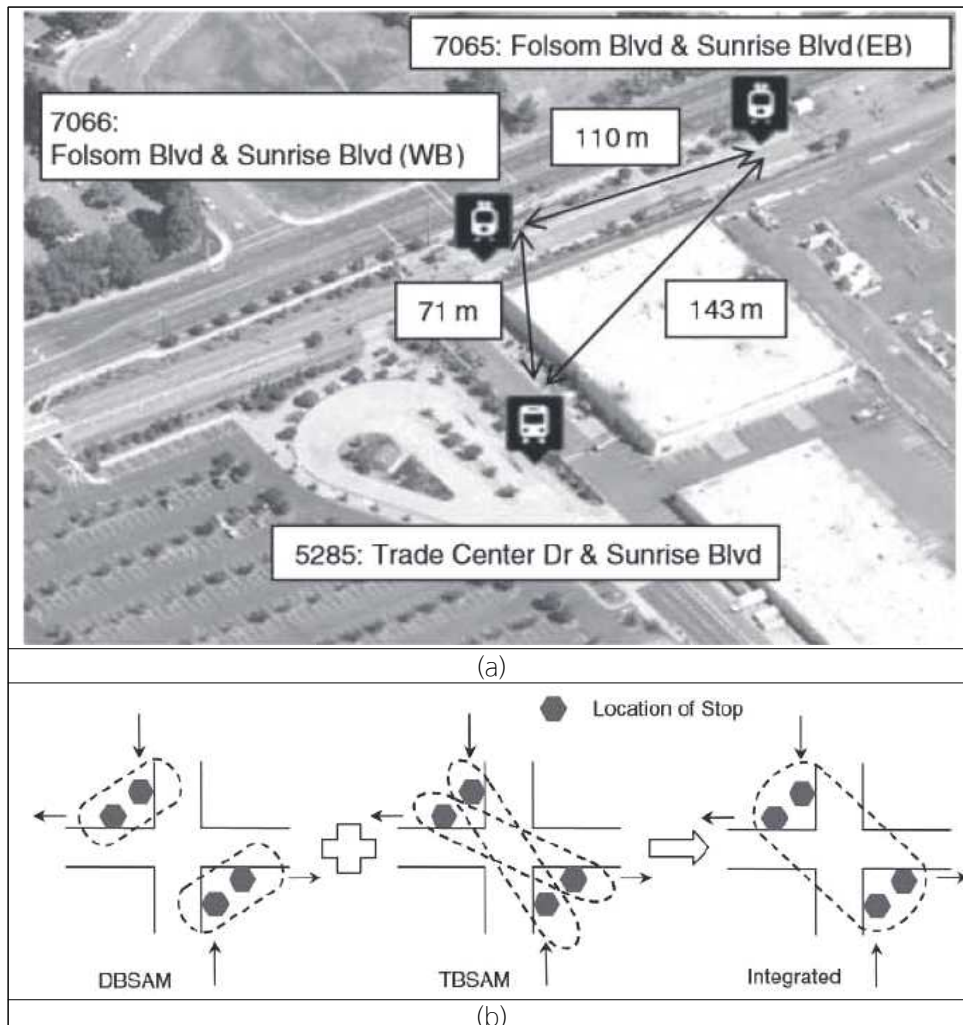
첫째, DBSAM은 개별 정류장 간 거리를 기준으로 정류장 통합이 이루어진다. 이를 위해 일정 거리를 임계값으로 설정하여 임계값 이내 정류장들을 단일 정류장으로 간주한다. Lee는 도심과 교외 지역, 동일 방향과 대향 방향 등에 따라 기준거리의 차이가 있을 수 있음을 지적한다.

둘째, TBSAM은 정류장 명칭 기준 통합모형으로 위치가 인접한 정류장들의 명칭이 유사하거나 동일한 경우에 단일 정류장으로 간주한다.

셋째, CBSAM은 정류장과 인근 토지이용 간의 관계가 정류장 통합 여부를 판단하는 기준으로 설정되는 경우로 통합의 정도가 가장 광범위하다고 할 수 있다.

<그림 3-1> (a)에는 SAM에 의한 공간적으로 분리된 3개 정류장 통합과정 예시가 제시되어 있다. 우선 100m를 거리 기준의 임계값으로 가정하여 임계거리 100m 내에 위치한 7066 정류장과 5285 정류장을 통합하고, 다음으로 임계거리는 초과되었으나 명칭이 동일하다고 판단된 7066 정류장과 7065 정류장을 통합한다. 마지막으로 3개 정류장 모두가 동일한 시설인 Trade Center에 접근성을 제공할 한다는 점을 고려하여 3개

정류장을 통합된 단일 정류장으로 간주한다.



자료 : Lee, S. G., Hickman M, Tong D, "Stop Aggregation Model Development and Application", Transportation Research Board, No.2276, (2012), p.41.

<그림 3-1> Integrated SAM: (a) Sunrise park-and-ride in Sacramento, California, and (b) conceptual approach of an integrated model.

그러나, Lee(2012)는 거리 기준에 따라 통합이 불가능한 거리상에 동일 명칭 정류장들이 존재하는 경우도 있고, 매우 인접한 정류장들의 명칭이 상이한 경우도 있으며, 거리와 명칭에 상관없이 주변 입지와 여건에 따라 동일한 정류장으로 고려하여야 하는 경우도 있다고 지적하면서

DBSAM, TBSAM 및 CBSAM를 모두 동시에 고려한 Integrated 방법도 동시에 제시한다. <그림 3-1> (b)에 4개 정류장이 DBSAM과 TBSAM을 종합적으로 고려하여 Integrated 방법에 의한 1개 정류장으로 통합된 예시가 제시되어 있다.

2. 통행사슬 수리 모형 구축

통행사슬 모형 적용 연구는 교통카드 자료에 나타난 각 이용자의 통행 간 시공간적 관계로부터 연속 통행 및 최종 통행 등 통행 형태로 구분하여 통행사슬을 구성한 후 각 승차 통행의 하차지점을 추정한다. 연구에 따라 시공간적 관계와 통행패턴에 관한 가정에 다소의 차이가 있으나, 본 연구는 대표적으로 적용되어온 아래 2가지 통행패턴에 관한 가정에 현실적 한계를 고려하여 통행사슬을 구성한 후 하차지점을 추정한다.

첫째, 두 번의 연속된 승차 통행 중 선(先) 통행의 하차 정류장은 후(後) 통행 승차 정류장의 도보권 내에 위치한 선 통행 노선의 통과 정류장들 중 가장 인접한 정류장이라는 가정과,

둘째, 하루 최종 승차 통행의 하차 정류장은 최초 통행 승차 정류장의 도보권 내에 위치한 최종 통행의 노선이 통과하는 정류장들 중 가장 인접한 정류장 된다는 가정이다.

본 연구는 교통카드이용자의 승차 통행(수단통행)에 대한 하차 정류장 추정을 목적으로 하며, 2가지 가정을 통행사슬 Rule 1과 통행사슬 Rule 2로 명명하여 수리적인 모형식으로 표현한다. 모형식은 교통카드이용자, 통행 순서(시간), 승차 및 하차, 탑승 노선, 각 노선의 정류장 등을 변수로 구성하며, 각 변수 기호를 아래와 같이 정의한다.

c : 교통카드이용자(Unique card ID) ($c=1\cdots n$)

i : 통행 순서 ($i=1\cdots\ell$)

a : 하차

b : 승차

r : 노선 (Route, Line)

m : 정류장 순서 ($m=1\cdots k$)

S : 정류장

$R(r)$: 노선 r 의 순서화된 정류장 집합

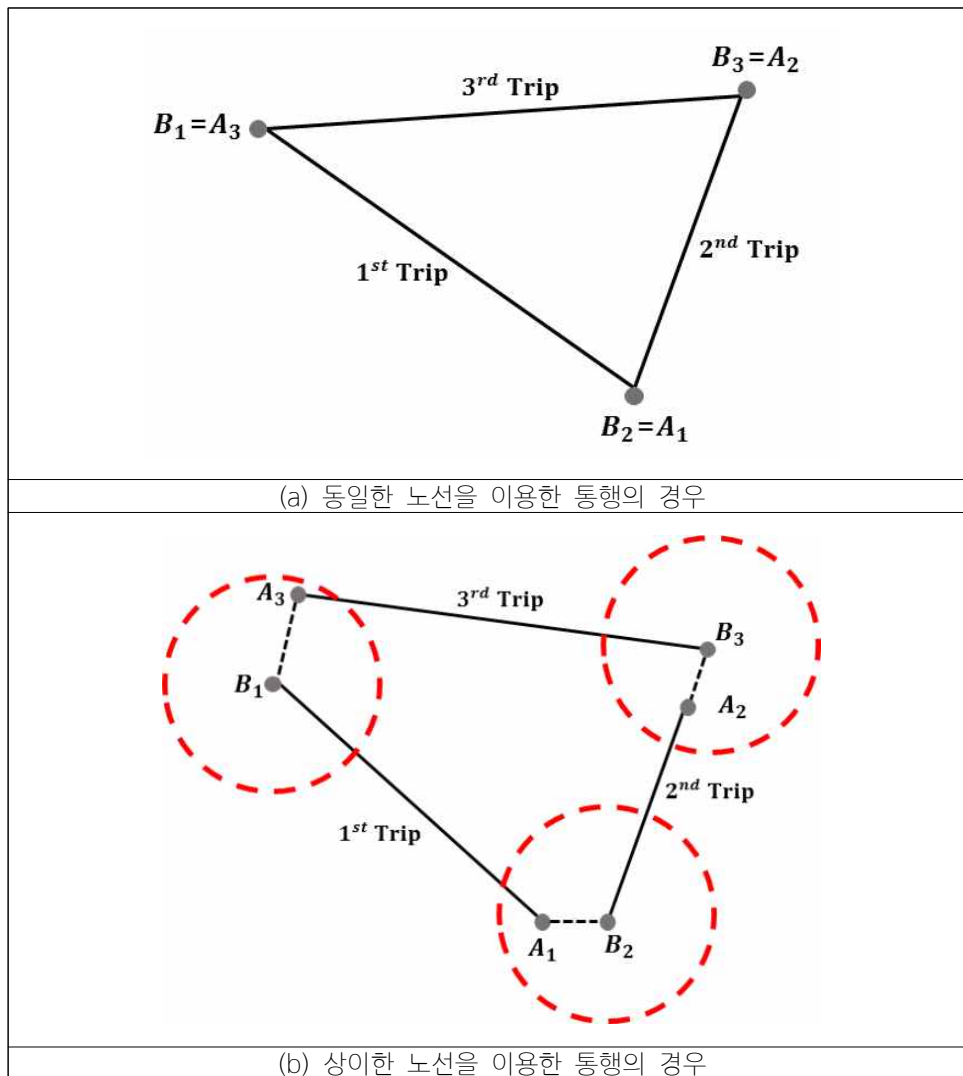
각 교통카드이용자가 1개의 교통카드만을 사용한다고 가정하면 교통카드 ID로 이용자 구분이 가능하며, n 은 총 이용자 수 혹은 이용된 교통카드 개수를 의미한다. c_i 는 교통카드이용자 혹은 교통카드 c 의 하루 통행 중 i 번째 통행을 나타내며, 마지막 통행은 c_ℓ 로 표현된다. 기점에서 종점을 거쳐 다시 기점으로 돌아오는 정류장까지 k 개의 정류장으로 구성된 노선 r 의 순서화된 정류장 집합인 $R(r)$ 은 정류장 순서 m 을 고려하여 $R(r) = \{S^{r_m} : S^{r_1}, S^{r_2}, \dots, S^{r_k}\}$ 로 표현된다. 여기서 S^{r_m} 은 노선 r 의 기점부터 종점을 거쳐 다시 기점으로 돌아오기까지의 정류장 중 m 번째 정류장이다. 마지막으로 $S_{c_i}(b)$ 는 교통카드이용자 c 의 i 번째 통행 승차 정류장을 나타낸다.

2.1 통행사슬 Rule 1

통행사슬 Rule 1은 두 번의 연속된 승차 통행 중 선(先) 승차 통행의 하차 정류장을 추정하기 위한 가정으로 분석의 추정률과 정확성 제고를 위해 연속된 2통행의 이용 노선이 동일한 경우와 상이한 경우로 구분한다.

첫째, 동일한 노선인 경우의 통행 간의 관계가 <그림 3-2> (a)에 도식화되어 있다. 특정 이용자가 세 번의 연속된 통행을 모두 동일 노선을 이용한 경우 첫 번째 승차 통행의 하차 정류장(A_1)은 두 번째 통행의 승차 정류장(B_2)으로, 두 번째 통행의 하차 정류장(A_2)은 세 번째 통행

의 승차 정류장(B_3)이 된다. 즉, 동일한 노선을 이용하는 경우 후(後) 통행 승차 정류장이 선(先) 통행 하차 정류장이 되는 이유는 후(後) 통행 승차 정류장이 선(先) 통행 노선의 정류장들 가장 인접한 거리의 정류장이라는 가정에 의해 인접거리가 0이 되기 때문이다.



<그림 3-2> Rule 1 : 하차 정류장 추정 모형

둘째, <그림 3-2> (b)는 각 통행이 상이한 노선을 이용할 경우로서, 첫 번째 통행의 하차 정류장(A_1)은 첫 번째 통행 노선의 정류장 중 두 번째 통행 승차 정류장(B_2)과 도보권 내 존재하는 가장 인접한 정류장이 된다. 따라서 그림상에 표현된 두 번째 및 세 번째 통행의 하차 정류장은 A_2 와 A_3 로 추정된다.

본 연구에서는 도보권을 버퍼 존 개념으로 설명하였다. 버퍼 존의 크기는 보통 시간 혹은 거리 단위로 정의되며, 일반적으로 도보 시간 5분 내지 10분 혹은 거리 400m 혹은 800m를 기준으로 설정한다. 만약, 다수의 정류장이 버퍼 존 내(즉, 정류장 B_2 의 버퍼 존)에 존재하는 경우 B_2 와 최단 거리인 대상 노선의 정류장 A_1 을 B_1 의 하차 정류장으로 추정하며, 만약, 버퍼 존 내에 대상 노선의 정류장이 존재하지 않는 경우 추정이 불가능한 것으로 판단되도록 모형을 구성하였다.

위 도식화된 통행사슬 모형의 Rule 1을 수리적 모형식으로 표현하면 아래와 같다.

첫째, 연속 통행의 노선이 동일한 경우, 후(後) 통행 승차 정류장이 선(先) 통행의 하차 정류장이 되기 때문에 버퍼 존 개념의 적용은 무의미하다. 즉, $r_{c_i} = r_{c_{i+1}}$ 인 경우, 이용자 c 의 선(先) 통행 추정 하차 정류장은 후(後) 통행의 승차 정류장이 되며, 이를 식으로 표현하면 아래와 같다.

$$\hat{S}_{c_i}(a) = S_{c_{i+1}}(b) \quad (1)$$

둘째, 연속 통행 노선이 상이한 경우는 버퍼 존 내에 선(先) 통행 노선이 통과하는 정류장 중 후(後) 통행 승차 정류장과 최단 거리 정류장을 찾아야 한다. $d(S', R(r))$ 를 노선 r 이 통과하는 순서화된 정류장 집합인 $R(r)$ 과 특정 정류장 S' 와의 다차원 공간에서 두 점(p, q) 간의 최단

직선거리인 유클리디언 거리(Euclidean Distance)의 집합이라 하고, 노선 r 의 정류장 중 특정 정류장 S' 와 최단 거리에 위치한 정류장을 $S(\min d(S', R(r)))$ 라 하면, 이용자 c 의 선(先) 통행 추정 하차 정류장은 아래 식으로 표현된다.

$$\widehat{S}_{c_i}(a) = S(\min d(S_{c_{i+1}}(b), R_{c_i}(r)), s.t. \min d \leq d_\chi \quad (2)$$

여기서 d_χ 는 추정을 위해 설정된 버퍼 존의 최대 거리이며, 제약 조건식은 최단 거리 정류장이 버퍼 존 내에 존재해야 추정이 가능하다는 조건 충족을 위한 것이다.

2.2 통행사슬 Rule 2

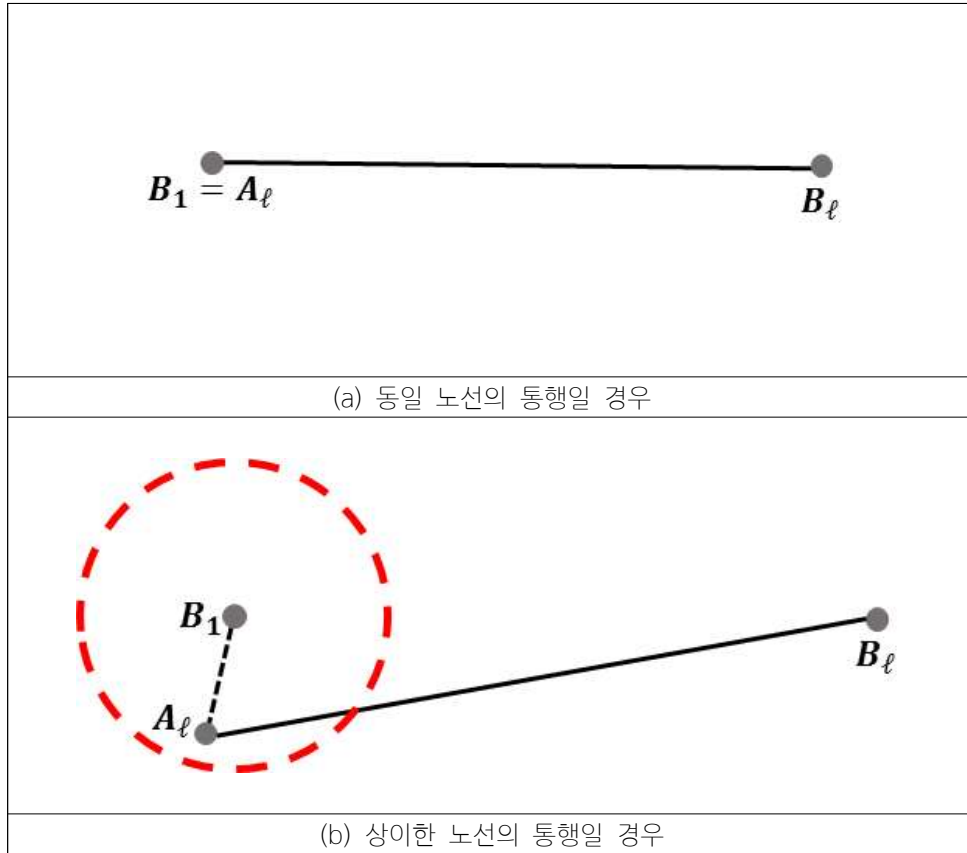
하루 최종 통행의 하차 정류장을 추정하기 위한 가정인 통행사슬 Rule 2는 Rule 1과 마찬가지로 최초 통행과 최종 통행의 노선이 동일한 노선인 경우와 상이한 경우로 나누어 동일한 논리로 설명한다. 즉, 최초와 최종 통행이 동일 노선에 의해 이루어지는 경우 최종 통행의 하차 정류장 (A_ℓ)은 최초 통행의 승차 정류장(B_1)이 되며, 상이한 노선인 경우 최종 통행의 하차 정류장은 최종 통행 노선의 통과 정류장 중 최초 통행의 승차 정류장과 가장 근접한 정류장으로 추정된다. <그림 3-3>의 (a)와 (b)에 동일노선과 상이한 노선인 경우의 통행 관계가 도식화되어 있다.

Rule 2를 수리 모형식으로 표현하면 아래와 같다.

첫째, 버퍼 존 적용이 무의미한 최초 통행과 최종 통행의 노선이 동일한 경우, 즉, $r_{c_1} = r_{c_\ell}$ 인 경우, 이용자 c 의 하루 최종 통행의 추정 하차 정류장은 아래와 같이 표현된다.

$$\widehat{S}_{c_\ell}(a) = S_{c_1}(b) \quad (3)$$

d_χ 는 추정을 위해 설정된 버퍼 존의 최대 거리임을 이미 언급된 바 있다.



<그림 3-3> Rule 2 : 하차 정류장 추정 모형

최초 통행과 최종 통행의 노선이 상이한 경우, Rule 1과 동일한 논리로 이용자 c 의 버퍼 존 내에 최종 통행 노선이 통과하는 정류장 중 최초 통행 승차 정류장과 최단 거리인 정류장이 추정 정류장이 된다. 이 경우 이용자 c 의 추정 하차 정류장은 아래 식으로 표현할 수 있다.

$$\widehat{S}_{c_\ell}(a) = S(\min d(S_{c_1}(b), R_{c_\ell}(r)), s.t. \min d \leq d_\chi) \quad (4)$$

3. 추정률 모형식

하차 정보가 부재한 통행 자료로부터 하차 정보 추정 시 적용된 가정과 버퍼 존을 포함한 통행에 관한 통행거리 제약조건에 따라 추정이 불가능한 통행이 불가피하게 발생할 수 있다. 그간 추정 불가능한 통행 규모의 최소화를 통한 추정률 제고를 목표로 버퍼 존 확대 혹은 가정의 다양화를 위한 노력이 진행되어 왔다. 이는 추정 불가 통행 규모가 클수록 통행사슬 모형의 적용성과 결과에 대한 신뢰성 저하 문제가 발생할 수 있기 때문이다.

대상 총 승차 통행 대비 추정된 하차 통행의 비율로 정의될 수 있는 추정률은 적용된 가정뿐 아니라, 수집 정보 내용 및 속성, 대상 지역 특성, 이용자 통행패턴, 시간대, 수단 구성 및 노선 구조 등과 같은 대중교통 시스템 특성에 의해서도 영향을 받을 수 있다. 실제 통행사슬 모형을 적용한 연구들 간에도 추정률에 상당한 차이를 보이고 있다. 예를 들면, 도시철도와 버스 이용자 교통카드 자료를 분석한 Zhao et al.(2007)는 71.2%의 추정률을, Munizaga et al.(2012)의 경우 월별 도시철도와 버스 이용자 자료를 분석하여 2009년 3월은 80.77%, 2010년 6월은 83.01%로 각각 차이가 있음을 보였다. 또한, 버스 교통카드 자료를 분석한 Li et al.(2011)는 전일 자료로는 70%와 침두시 자료로는 80%의 추정률을 제시하여 시간대 별로 차이가 있을 수 있음을 지적하고 있다. 즉, 연구에 따라 보고된 추정률은 차이가 있으며, 개략적으로 60.7%~83.0% 사이의 추정률을 보고하고 있다.<표 2-1> 참조)

추정률이 시간대, 수단 구성, 수집 정보의 내용과 대상 지역/도시의 이용자의 통행패턴 등에 따라 가변적임을 인지하고, 본 연구는 추정률(ε)을 하차 정류장이 추정된 통행($\hat{S}_{c_i}(a)$)과 통행사슬 가정에 의해 하차 정류장 추정이 가능하다고 판단된 승차 통행($S_{c_i}(b)$) 총계와의 비율로 정의하여 식 (5)로 표현하였다.

$$\varepsilon = \frac{\sum_c \sum_i \hat{S}_{c_i}(a)}{\sum_c \sum_i S_{c_i}(b)} \quad (5)$$

여기서 ε 는 추정률이며, 분모는 추정이 가능한 승차 통행만을 포함한다. 즉, 하루 1 승차 통행 교통카드이용자 자료와 속성 자료 부재 및 오류 등 다양한 사유로 추정이 불가능한 승차 통행 자료는 제외된다. 여기서 선행연구의 대부분이 언급된 추정 불가능한 자료를 포함하여 추정률을 제시하고 있기 때문에 본 연구 결과와 비교를 위해서는 추정 불가 자료를 제외한 추정률로 산정하기 위한 보정 작업이 필요하다.

통행사슬 Rule 1과 Rule 2를 구분하여 추정률을 표현하면 식(6)과 식(7)로 각각 표현된다.

$$\varepsilon_1 = \frac{\sum_c \sum_i \hat{S}_{c_i}(a) - \sum_c \sum_\ell \hat{S}_{c_\ell}(a)}{\sum_c \sum_i S_{c_i}(b) - \sum_c \sum_\ell S_{c_\ell}(b)} \quad (6)$$

$$\varepsilon_2 = \frac{\sum_c \sum_\ell \hat{S}_{c_\ell}(a)}{\sum_c \sum_\ell S_{c_\ell}(b)} \quad (7)$$

ε_1 은 Rule 1, ε_2 는 Rule 2의 추정률을 각각 나타낸다. ε_2 는 하차 정류장 추정 대상인 최종 통행의 총합과 최종 통행 중 하차 정류장이 추정된 통행의 총합과의 비율이며, ε_1 은 전체 통행에서 최종 통행인 Rule 2의 값들을 분모와 분자 값에서 각각 제외한 값으로 산정된 추정률이다.

4. 일치도 모형

통행사슬 모형의 신뢰성은 추정된 하차 정류장이 얼마나 실제 하차 정류장과 일치하는지에 달려 있다. 만약 추정 하차 정보가 실제 하차 정보와 큰 차이가 있다면 통행사슬 모형은 신뢰할 수 없는 방법론으로 전락하게 된다. 무엇보다도 일치도 검증의 전제 조건은 실제 하차 정보의 존재 유무이기 때문에, 추정 값과 실제 값과의 비교를 통하여 차이를 확인하는 과정을 거치는 일치도 검증은 제한적으로 이루어져 왔다. 제 2장에서 언급하였듯이 대부분의 선행연구가 기존 통행실태 자료를 활용하거나 설문조사 혹은 현장 혹은 실태조사 자료를 활용하여 동일한 날짜의 교통카드 자료와 비교하거나, APC(Automatic Passenger Counter)와 버스정보시스템(Bus Information System) 자료 등을 활용하여 90%에 가까운 일치도를 제시하기도 하였다. 그러나 비교 자료 자체의 신뢰성 문제와 소규모 수집 자료의 한계 등의 문제로 신뢰할 만한 검증 연구로 간주하기에는 한계가 있다. 최근 Alsger(2015, 2016)는 하차 정보가 알려진 자료를 활용하여 통행사슬 모형의 일치도 검증을 시도하였고, 일치도는 버퍼 존의 시간 혹은 거리의 가정뿐 아니라 추정 값의 수준(정류장(Stop), 존(Zone) 등)에 따라 가변적임을 지적하였다.

본 연구의 일치도는 실제 하차 정류장 정보가 존재하는 승차 통행 중 추정 하차 정류장과 실제 하차 정류장이 일치하는 통행과의 비율로 표현한다. 추정 하차 정류장과 실제 하차 정류장이 일치하는 통행의 하차 정류장을 아래와 같이 $\ddot{S}_{c_i}(a)$ 로 정의하면,

$$\hat{S}_{c_i}(a) = S_{c_i}^{AD}(a) \rightarrow \ddot{S}_{c_i}(a) \quad (8)$$

여기서 AD 는 실제 하차를 의미하며, $S_{c_i}^{AD}(a)$ 는 이용자 c 의 i 번째 통행의 실제 하차 정류장을 말한다.

일치 비율 계산식인 일치도는 실제 하차 정류장과 추정 하차 정류장이 동일한 통행($\ddot{S}_{c_i}(a)$)과 실제 하차 정류장 정보가 존재하는 통행 중 하차 정류장이 추정된 승차 통행($\hat{S}_{c_i}^{AD}(a)$)의 총계와의 비율로서 식 (9)로 표현된다.

$$\rho = \frac{\sum_c \sum_i \ddot{S}_{c_i}(a)}{\sum_c \sum_i \hat{S}_{c_i}^{AD}(a)} \quad (9)$$

여기서, ρ 는 일치도를 나타낸다.

위 식을 추정률과 동일한 논리를 적용하여 통행사슬 Rule 1과 Rule 2에 따라 구분하여 일치도를 식으로 나타내면 식 (10)과 식 (11)로 각각 표현된다.

$$\rho_1 = \frac{\sum_c \sum_i \ddot{S}_{c_i}(a) - \sum_c \sum_\ell \ddot{S}_{c_\ell}(a)}{\sum_c \sum_i \hat{S}_{c_i}^{AD}(a) - \sum_c \sum_\ell \hat{S}_{c_\ell}^{AD}(a)} \quad (10)$$

$$\rho_2 = \frac{\sum_c \sum_\ell \ddot{S}_{c_\ell}(a)}{\sum_c \sum_\ell \hat{S}_{c_\ell}^{AD}(a)} \quad (11)$$

여기서, ρ_1 은 Rule 1, ρ_2 는 Rule 2의 일치도를 각각 나타낸다.

제 4 장 자료 수집 및 가공

본 연구는 부산시 대중교통 시스템을 사례로 분석 모형 적용을 시도하였다. 도시철도, 시내버스 및 마을버스 수단으로 구성된 부산시 대중교통 시스템은 하루 평균 2,700천 통행을 처리하며, 수단분담률은 41.0%에 이르며, 도심 내부뿐만 아니라 인근 지역과 연계 서비스를 제공하는 부산시의 대규모 핵심 요소이다. 대중교통이용자의 교통카드 이용률은 95%이며, 무료 환승 제도로 인해 하차 정류장 정보가 생성되는 통행 비율인 환승률은 2016년 기준 30.5%로 나타났다(부산광역시, 2017).

본 장에서는 연구 목적 달성을 위해 수집된 교통카드 자료 및 필요 자료의 정리 및 가공 과정을 기술한다. 수집 자료들은 분석 모형에 적합하도록 구성되어야 하기 때문에 수집된 원자료의 속성 검토를 통해 모형식에 요구되는 정보가 도출될 수 있도록 정리 및 가공을 시도하였다.

구체적으로, 우선 정류장 수준(Stop-level) 하차 정보 추출을 위해 정류장 통합모형(SAM)을 적용하여 모든 대중교통수단의 공간적으로 산재한 정류장을 통합한 후, 수집된 교통카드 원자료 속성을 검토하여 분석에 필요한 속성을 추출하고, 다음으로 추출된 속성을 정리/가공하여 모형에 적합하고 분석에 용이한 자료 형태로 가공하였다.

1. 정류장 통합

제 3 장에서 언급된 SAM을 적용하기 위해 부산시 대중교통수단 - 즉, 도시철도, 시내버스 및 마을버스-의 모든 노선별 정류장 정보를 수집하였다. 수집된 노선별 정류장 정보는 정류장 순서, 코드, 명칭 및 좌표 등 분석에 필요한 핵심 속성을 포함하고 있다.

정류장 특성이 버스와 다소 상이한 108개 도시철도역을 제외하고 자료 수집 일자인 2016년 6월 14일 기준 정류장 코드로 구분되는 시내버스와

마을버스 정류장 개수는 14,399개인 반면, 정류장 명칭으로 구분할 시 정류장 개수는 4,552개로 감소한다. 마지막으로 본 연구의 통합 대상인 좌표로 구분되는 위치별 정류장 개수는 총 8,025개가 존재하는 것으로 파악되었다. 수단별/속성별 구분된 정류장 개수가 <표 4-1>에 제시되어 있다.

<표 4-1> 부산시 교통수단별 정류장 개수

속성	시내버스	마을버스	도시철도	총계
정류장 코드	9,454개	4,945개	108개	14,507개
정류장 명칭	2,041개	2,511개		4,660개
정류장 위치(좌표)	3,969개	4,056개		8,133개

본 연구는 제 3 장의 Lee et al.(2012)가 제시한 거리, 명칭 및 인접 시설 등의 통합기준과 적용 순서를 준용하되 정류장별 통과 노선 일치 여부 및 정류장 위치의 여건 등을 고려하여 아래 원칙에 따라 통합을 시도하였다.

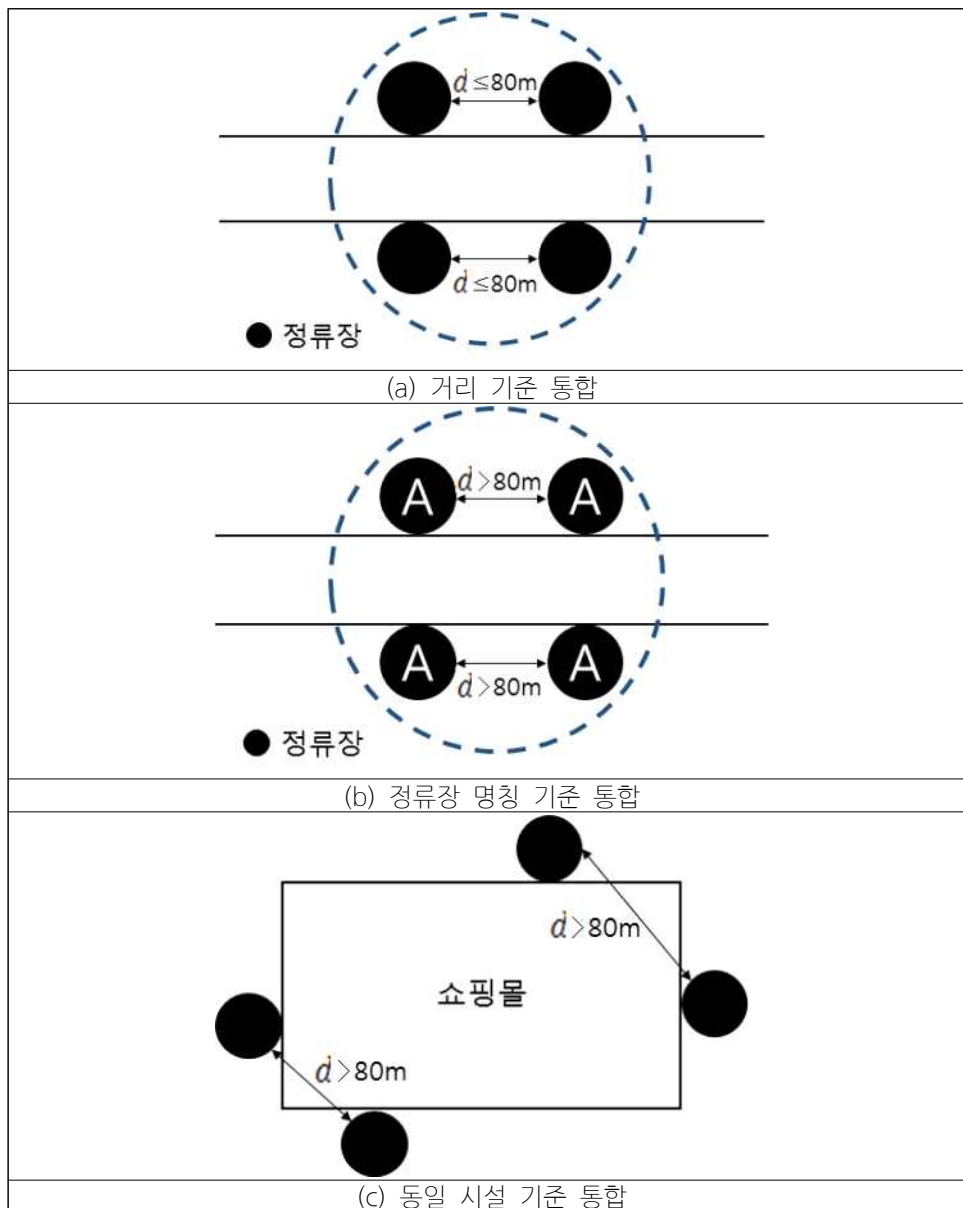
첫째, 직선거리 통합기준의 임계값을 부산시 평균 버스 정류장 간 거리인 약 240m의 1/3값인 80m로 설정하였다. 즉, 직선 80m 이내 존재하는 동일 방향 정류장은 명칭에 상관없이 모두 단일 정류장으로 간주하였고, 반면 대향 정류장은 필요한 경우 도로폭을 제외한 직선거리 80m로 가정하였다. 단 통합 대상 정류장과 동일 방향 80m 이내에 동일한 노선이 통과하는 정류장이 존재하는 경우 80m 이내임에도 독립 정류장으로 간주하여 통합 대상에서 제외하였다.

둘째, 거리 기준 임계값은 초과하나 동일 혹은 유사 명칭의 정류장은 통과 노선이 중복되지 않는 경우 단일 정류장으로 통합하였다. 즉, 80m 이상 거리인 동일 혹은 대향 방향 정류장들의 명칭이 동일하거나 유사하며, 동일 노선이 통과하지 않는 경우 단일 정류장으로 통합하였다.

셋째, 동일한 시설에 인접한 정류장들은 거리 및 명칭에 상관없이 단

일 정류장으로 통합하였다.

<그림 4-1>에 위 기술된 3가지 원칙이 도시화되어 있다.



<그림 4-1> 통합 정류장 기준

위 기준에 따라 통합된 정류장에 대해서는 정류장 명칭 수정 작업을

진행하였다. 기존 명칭 중 대표적 명칭을 부여하거나, 인근의 Landmark, 지역적 특성을 고려한 새로운 명칭을 부여하기도 하였다. <그림 4-2>에 정류장 통합 및 명칭 도출 사례가 제시되어 있다.



<그림 4-2> 정류장 통합 및 명칭 도출 과정 예시

<그림 4-2> (a)는 방향별로 '개금역'과 '개금초등학교' 2개의 정류장으로 총 4개의 정류장이 존재하고 동일 방향 정류장의 통과 노선은 상이하다. 또한, 동일 방향 2개의 정류장은 각각 80m 이내 거리이나, 대향 정류장 간은 80m 이상의 거리로 위치된 경우이다. 위 4개 정류장을 통합하는 과정은 다음과 같다.

첫째, 임계거리 80m 이내인 동일 방향의 2개 정류장을 1개 정류장으로 통합하고 2개의 상이한 명칭을 모두 표현할 수 있도록 새 명칭인 '개금초(개금역)'으로 명칭을 통합 변경한다.

둘째, 위 과정에서 동일한 명칭으로 변경된 대향 방향 2개의 정류장을 명칭 기준으로 통합한다.

즉, 거리와 명칭 기준에 의해 4개 정류장의 정류장을 1개 정류장으로 통합하였으며, 통합된 정류장의 위치는 4개 정류장 중 무게중심에 해당되는 위치의 좌표를 계산하여 설정하였다.

<그림 4-3>은 통합의 세 번째 기준 사례를 도시하고 있다.

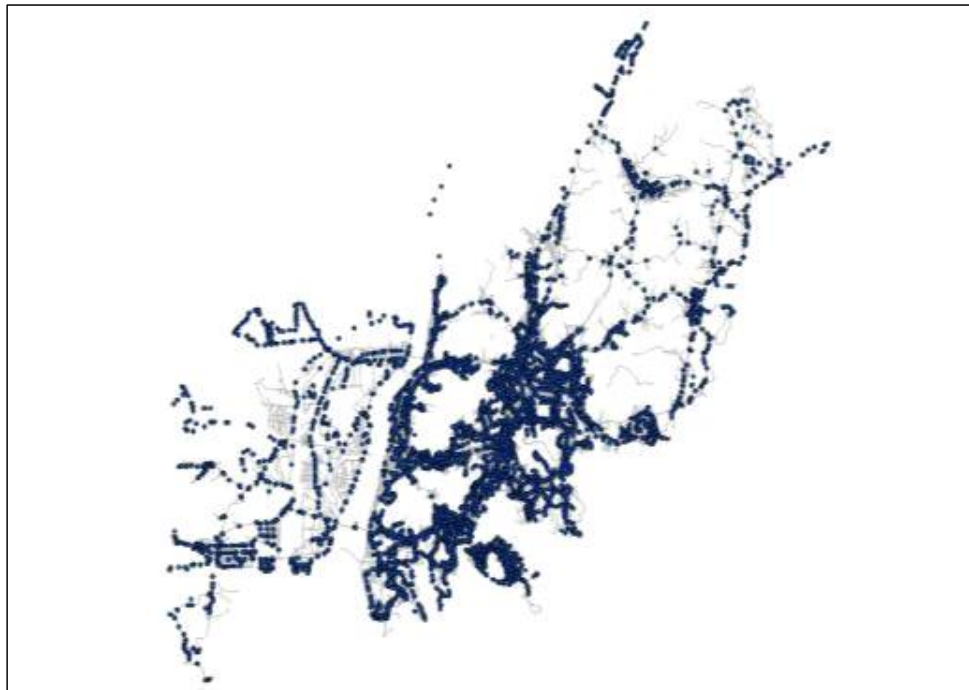


<그림 4-3> 통합 정류장 도출 과정 예시

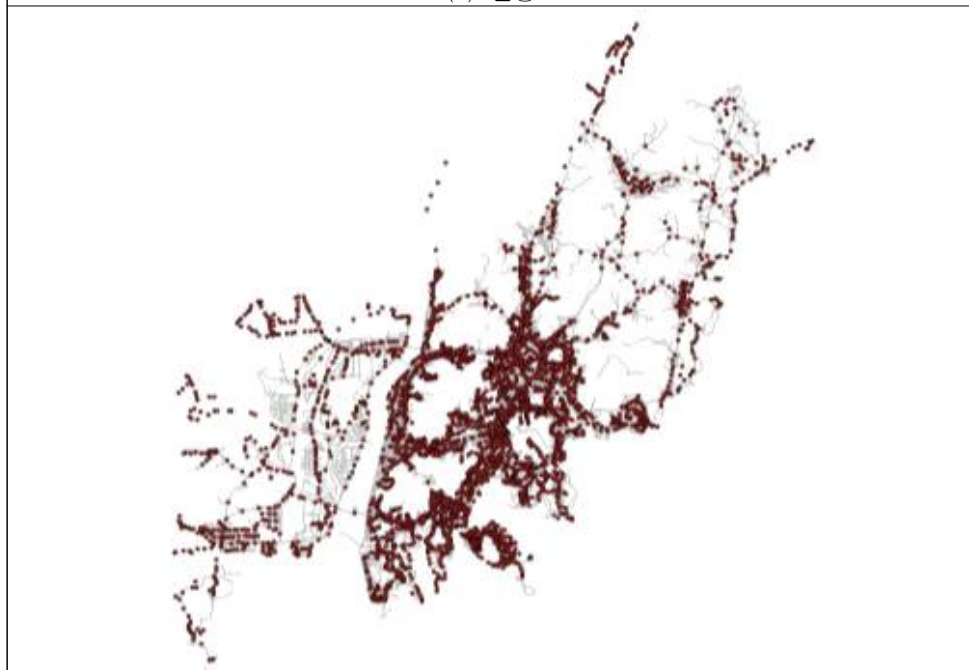
사상구에 위치한 판매시설을 포함하는 르네시페 쇼핑몰에 인접하여 위치한 3개의 시내 및 마을버스 정류장은 모두 동일 쇼핑몰에 접근성을 제공하는 하나의 정류장으로 간주할 수 있기 때문에 거리 및 명칭 기준에 상관없이 ‘르네시페’ 명칭으로 통합하였다.

대부분의 정류장들이 제시된 기준에 의해 통합이 가능하였으나, 현실적 여건으로 인해 기준에 의한 통합이 불가능하거나 용이하지 않은 경우가 발생하였고, 이를 해결하기 위해 주변 여건, 정류장 통과 노선, 실제 접근 거리, 도심 혹은 외곽 등 다양한 상황을 고려하여 통합을 시도하였다. 통합 결과 총 8,025개의 좌표 위치가 도시철도역 108개를 포함하여 총 3,552개의 상이한 정류장으로 축소 통합되었다.

<그림 4-4> (a)와 (b)에 부산시의 좌표로 구분된 모든 버스 및 마을버스 8,025개 정류장과 이를 위 기준에 따라 통합한 3,552개 정류장의 위치가 각각 표기되어 있다. 또한, <그림 4-5> (a)와 (b)에 해운대 신시가지 지역의 좌표 정류장과 통합 정류장의 위치가 표기되어 있으며, 이를 통해 통합 정류장이 기존 좌표 정류장의 접근 권역을 모두 포괄하고 있음을 알 수 있다.

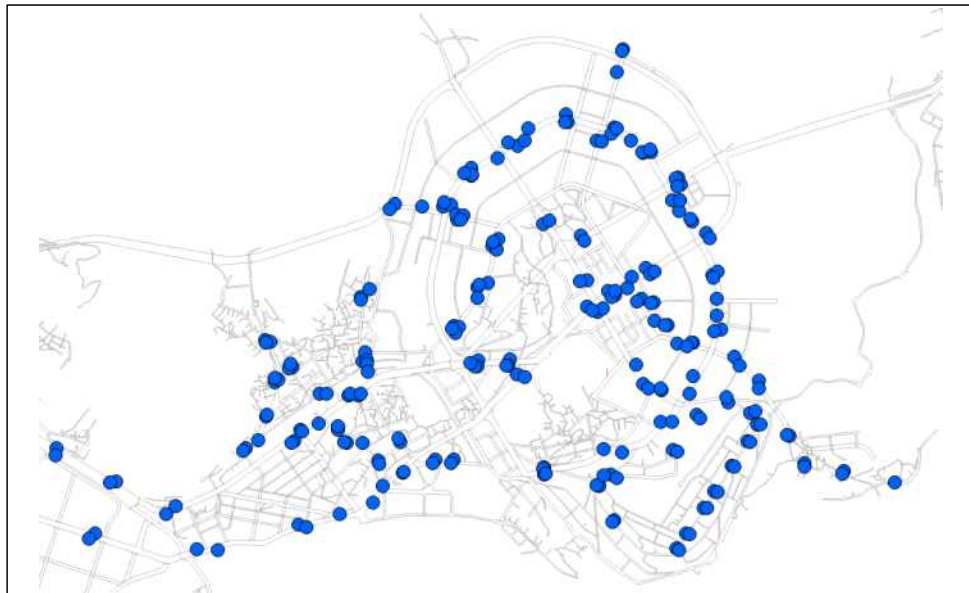


(a) 현황



(b) 통합 정류장

<그림 4-4> 현황 및 통합 정류장 비교



(a) 현황



(b) 통합 정류장

<그림 4-5> 해운대 신도시 지역 좌표 및 통합 정류장

2. 교통카드 자료 수집 및 가공

본 연구에서는 2016년 6월 14일 생성된 부산지역내 대중교통이용자의 교통카드 자료를 수집하여 구축된 모형에 적합하도록 데이터 정리 및 가공을 시도하였다. 부산시 교통카드시스템의 개략적인 내용과 수집된 부산시 교통카드 자료의 정리 및 가공 과정은 아래와 같다.

2.1 부산시 교통카드 개요

요금 지불의 편의성 제고와 요금 수집 및 관리를 목적으로 1998년 도입된 부산시 교통카드시스템은 교통카드이용자의 모든 통행 및 거래 기록을 자동적으로 수집하고 있다. 부산시 대중교통이용자가 사용하는 교통카드의 종류는 크게 충전이 요구되는 선불 교통카드와 충전이 필요하지 않은 후불 교통카드로 구분할 수 있다. 선불 교통카드로는 현재 단종된 하나로 카드와 통용 중인 (주)Mybi 카드(마이비 카드), 그리고 전국 호환 카드에 해당되는 T-money 등을 포함하며, 후불 교통카드는 교통카드 칩이 내장된 거의 모든 종류의 신용카드가 해당된다.

교통카드 종류에 따라 생성되는 속성 정보에 다소의 차이가 있다. 예를 들면, 마이비 카드의 경우 카드 거래 시 카드 ID를 포함 총 98개 속성 정보가 생성되며, 하나로 카드는 총 78개 속성 정보가 생성된다. <부록 1>에 부산시 마이비 카드 자료 총 98개 속성이 정리되어 있다. 대부분의 속성 정보는 교통카드시스템의 핵심 목적인 요금 수집과 수입금의 관리를 위해 활용되고 있다.

거리비례제를 시행하는 서울시와 달리 부산시는 2008년 마을버스를 제외한 수단 간 환승할인제를 시작으로 2011년 마을버스를 포함하여 수단 간 광역 환승할인제를 시행하였고, 2015년 모든 수단 간 환승 시 무료 환승제를 도입하였다. 이에 따라 이용자는 환승 시 단말기에 태그를 하고 하차 후 30분 이내 수단에 상관없이 동일 노선을 제외한 수단으로 갈

아탈 경우 무료 혜택이 제공되며 총 2회까지 환승이 허용된다.

무료 환승제로 인해 교통카드 이용 시 환승 이용자의 실제 하차 정류장 정보가 생성된다. 2016년 말 기준 부산시는 총 승차 통행의 30.5%가 환승을 하는 것으로 파악되었다(부산광역시, 2017). 즉, 부산시 교통카드 시스템은 총 통행의 30%에 해당되는 통행인의 실제 하차 정류장 정보를 제공하고 있다고 할 수 있다.

2.2 필요 속성 추출

수집된 교통카드 속성 자료를 검토하여 하차 정보 추정과 일치도 검증에 요구되는 필요 속성만을 추출하였다. <표 4-2>에 추출된 속성과 속성에 대한 설명이 제시되어 있다. 분석 모형식에 요구되는 추출 속성은 교통카드 번호, 차량 유형, 노선 코드, 승/하차 코드, 정류장/역 코드, 환승 유형과 거래 기록 날짜 및 시간으로 총 7개 속성이다.

<표 4-2> 추출 데이터 속성과 내용

속성	설명	예시
CARD_NO	교통카드 번호	234657020
VEHICLE_TYPE	차량 유형	0(시내버스), 1(도시철도), 2(마을버스)
EB_LANE_NO	노선 코드	26021008
ON_OFF_FLAG	승/하차 코드	0(승차), 1(하차)
ST	정류장/역 코드	2603445
HS_FLAG	환승 유형	0(비환승), 1(1환승), 2(2환승)
TRX_TIME	거래 기록 날짜&시간	20160614110324

CARD_NO는 이용자의 교통카드 식별 번호로서 신용카드의 경우 카드 번호(개인정보보호를 위해 변환하여 생성됨)가 여기에 해당된다. 대중교통이용자가 하나의 교통카드만을 사용한다고 가정하면 각 교통카드 번호

를 개별 대중교통이용자로 간주할 수 있다. VEHICLE_TYPE은 차량 유형을 의미하며, 이용자가 탑승한 시내버스, 도시철도 및 마을버스 수단을 구분할 수 있는 속성이다. EB_LANE_NO는 이용자 탑승 수단의 노선번호를, ON_OFF_FLAG로 승차 기록 혹은 하차 기록 여부 판단이 가능하며, ST는 이용자의 승차 혹은 하차 정류장의 코드를 나타낸다. HS_FLAG로 해당 이용자 통행의 환승 여부 파악이 가능하며, 첫 번째 환승 승차는 1로 두 번째 환승 승차는 2로 기록이 된다. 마지막으로 TRX_TIME는 거래 기록 날짜 및 시간을 나타내는 것으로 단말기에 교통카드를 태그 한 날짜와 시간을 초 단위까지 표현한다.

2.3 자료 정리 및 가공

추출된 속성 자료들을 우선 교통카드 번호로 정렬하여 전체 자료를 이용자별로 구분하였고, 다음으로 거래시간 순서로 정렬하여 각 이용자의 통행 순서(Trip-Sequence)를 파악하였다. 위 순서대로 정리된 자료 예시가 <표 4-3>에 제시되어있다.

<표 4-3> Raw Data의 자료정리 예시

Trip_ sequence	CARD_ NO	VEHICLE - TYPE	EB_LANE_NO	ST	ON_ OFF	HS_ FLAG	TRX_TIME
1-1	xxxx	1	00000001	0000125	0	0	20160614110324
1-2	xxxx	1	00000001	0000121	1	0	20160614111153
2-1	xxxx	0	26012002	2607663	0	1	20160614111820
2-2	xxxx	0	26012002	2607694	1	1	20160614120110
3	xxxx	2	26420001	2600787	0	2	20160614122132

교통카드 번호 xxxx인 교통카드이용자는 2016년 6월 14일 11시 03분 24초 당일 첫 번째 통행으로 도시철도 1호선 0000125 역에서 승차하여

11시 11분 53초에 0000121 역에서 하차하였고, 두 번째 통행은 시내버스 26012002번 노선의 2607663 정류장에서 11시 18분 20초에 1환승 승차를 하고 2607694 정류장에 12시 01분 10초에 하차하였다. 마지막 통행은 마을버스 노선 26420001의 정류장 2600787에서 12시 21분 32초에 2환승 승차로 이루어졌다. 본 교통카드 이용자는 도시철도 1통행, 시내버스 1통행, 마지막으로 마을버스 1통행, 총 3통행으로 하루를 마감하였다. 그중 하차 정류장이 기록된 통행은 첫 번째와 두 번째 통행이며, 하루의 마지막 통행은 승차 기록만 있을 뿐 하차 기록이 존재하지 않는다.

분석의 편의성과 용이성을 위해 Raw Data의 속성들을 변환하였다. <표 4-4>에 제시되어 있듯이 영문표기 속성을 한글 명칭으로 변환하였고, 숫자로 표기된 속성들의 코드를 실제 값을 부여하거나 이해가 용이한 표현으로 변환하였다. 예를 들면, 정류장 코드는 실제 명칭과 정류장 통합을 통해 도출된 통합 명칭으로 표현하였다.

<표 4-4> 데이터 가공 및 속성 변환

속성	기존 데이터	가공 데이터
교통카드 번호	xxxx	xxxx
차량 유형	0	시내버스
노선번호	26012002	86
정류장 명칭	2607663	양정
통합 명칭		양정(교차로)
승하차 구분	0	승차
환승 유형	1	1환승
거래시간	20160614111820	11:18:20

<표 4-3>에 제시된 자료를 변환된 속성으로 <표 4-5>에 표현할 경우가 제시되어 있다. 차량 유형 0은 시내버스, 1은 도시철도, 2는 마을버스로 표현하였고, 노선번호는 실제 노선번호로 정류장 코드는 정류장 명칭과

통합 명칭으로 변환하였다. 또한, 승하차 구분의 0은 승차 1은 하차, 환승 유형 0은 비환승, 1은 1환승 그리고 2는 2환승으로 속성을 표현하였다.

〈표 4-5〉 속성 변환 자료 예시

통행 순서	교통카드 번호	차량 유형	노선	정류장 명칭	통합 명칭	승하차 구분	환승 유형	거래시간
1-1	xxxx	도시 철도	1호선	동래역	동래역	승차	비환승	11:03:24
1-2	xxxx	도시 철도	1호선	양정역	양정역	하차	비환승	11:15:3
2-1	xxxx	시내 버스	86	양정	양정 (교차로)	승차	1환승	11:18:20
2-2	xxxx	시내 버스	86	국제시장	국제시장	하차	1환승	12:01:40
3	xxxx	마을 버스	중구1	부평시장	부평시장	승차	2환승	12:21:32

2.4 오류 자료

수집된 교통카드 원자료 중 오류로 인해 생성 자료로부터 정보 추출이 불가능한 경우가 발생하였다. 예를 들면, 정류장 코드와 같은 중요 속성 정보가 missing 되거나 zero로 표현되어 정보 파악이 불가능한 자료, 승차와 하차 정류장이 동일한 경우, 승차 통행 없이 하차 통행으로 시작하는 자료 그리고 시작 통행이 환승인 경우 등 다양한 사유로 분석 대상에서 제외되어야 할 자료가 다수 존재하였다.

이와 같은 자료는 속성 정보의 오류로 인해 하차 정류장 추정이 불가능한 자료이며, 특히 승하차 정류장이 동일한 자료의 경우 승차 자료는

분석에 사용은 가능하지만, 하차 정보는 신뢰할 수 없는 자료로서 분석에 부적절한 것으로 판단할 수 있다. 또한, 승차 통행 기록 없이 하차 통행 기록으로 시작되는 자료의 대부분은 광역 통행에 해당되는 것으로 판단되나, 승차지점을 알 수 없기 때문에 분석 대상에 제외하였다. 마찬가지로 논리로 환승으로 시작되는 통행 자료도 분석에서 제외하였다. <표 4-6>에 오류 자료의 사유가 요약 제시되어 있다.

<표 4-6> 오류 자료 사유

구분	오류 자료
1	속성 정보가 없는 통행 자료
2	승차와 하차 정류장이 동일한 자료(승차는 사용 가능)
3	승차 통행 없이 하차 통행으로 시작하는 자료
4	환승 순서 오류 자료 (ex) 선(先) 통행 2환승 → 후(後) 통행 1환승 기록 자료

3. 분석 대상 데이터

수집된 자료의 가공 및 오류 자료 제외 작업을 거쳐 최종적으로 분석 데이터를 확정된 결과가 <표 4-7>에 요약되어 있다.

<표 4-7> 분석 대상 데이터 요약

항목	데이터 수(통행)
총 기록	3,473,772
속성 정보가 없는 기록(오류 1)	11
승차 기록	2,309,893
하차 기록	1,163,770
승하차 정류장 동일 하차 기록 등(오류 2, 3, 4)	7,254
오류 2,3,4 제외 하차 기록	1,156,516
1통행	310,798
1통행 중 실제 하차 정류장 존재 기록	112,272
총 분석 대상 승차 기록	1,999,095
총 실제 하차 정류장 존재 기록	1,044,244

2016년 6월 14일 수집된 교통카드 자료의 승차 및 하차 기록은 총 3,473,772 통행으로 파악되었다. 총 기록으로부터 속성 정보 부재 자료 11개의 통행을 제외할 경우 승차 2,309,893통행 기록과 하차 1,163,770통행 기록이 남는다. 여기서 오류자료 2, 3 및 4 사유에 해당되는 7,254통행 하차 기록과 카드 번호 당 승차 혹은 승차와 하차만 존재하는 하루 1통행 기록 310,798통행을 모두 제외하면, 분석 대상 승차 기록은 1,999,095통행으로 파악되었다.

하루 1통행 기록이 전부 제외된 사유는 통행사슬 모형으로 하차 정류장 추정이 불가능하기 때문이며, 제외된 1통행 자료 중 하차 기록 112,272통행을 제외한 1,044,244통행 기록만이 실제 하차 정보를 제공하는 하차 기록에 해당되기 때문에 분석 대상 승차 통행 1,999,095통행 중 52.23%에 해당되는 통행이 실제 하차 정류장 정보가 알려져 있다고 간주하면 된다.

본 연구가 구축한 통행사슬 모형 적용은 승차 기록인 1,999,095통행을 대상으로 하차 정류장 추정이 이루어지며, 추정 하차 기록을 토대로 승차 통행에 대한 추정률을 산정하며, 실제 하차 정류장 통행 기록이 존재하는 승차 기록 (1,044,244통행) 중 추정 하차 기록이 존재하는 통행만을 대상으로 추정 하차 정류장과 실제 하차 정류장의 일치 여부인 일치도를 산정한다.

제 5 장 추정률과 일치도 분석

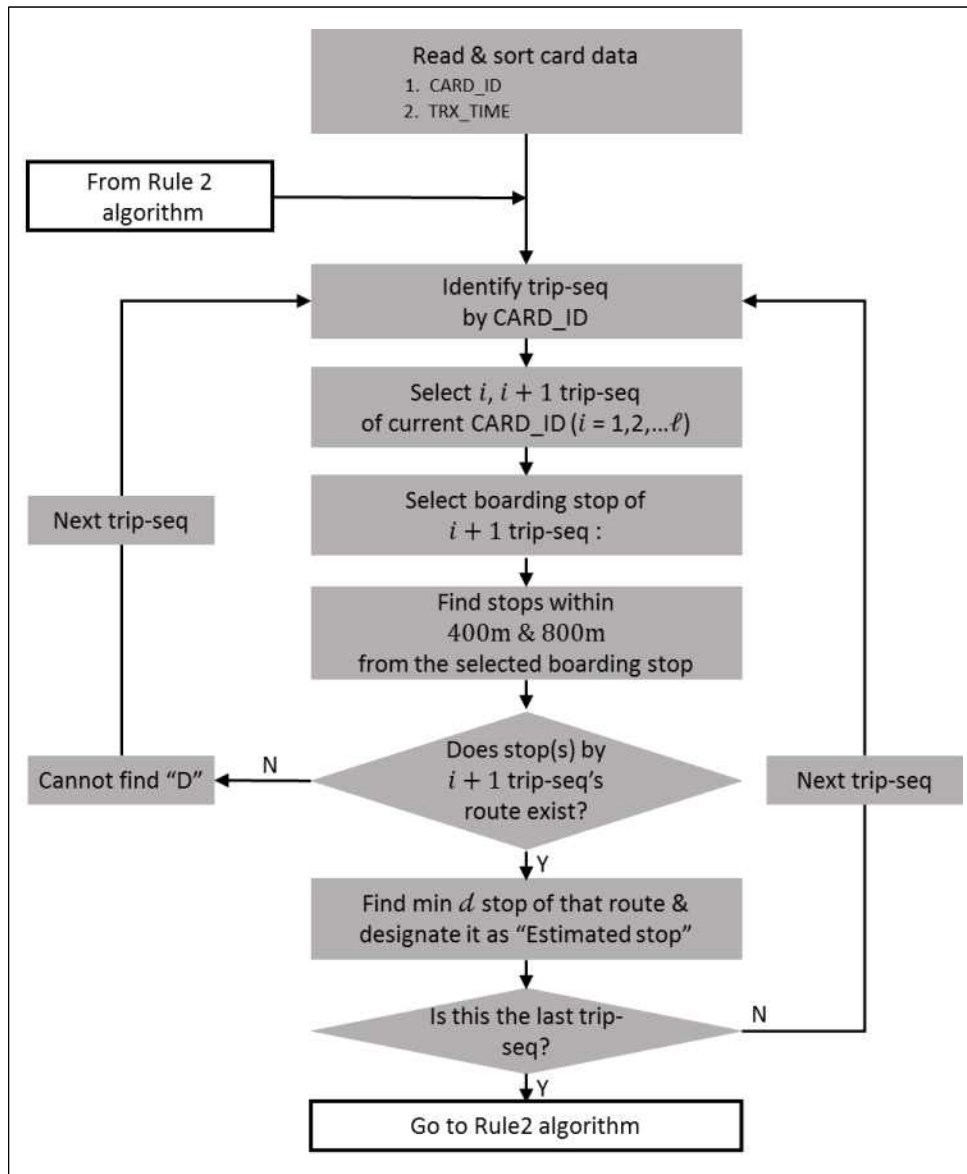
본 장에서는 통행사슬 모형 분석 모형식인 통행사슬 Rule 1과 통행사슬 Rule 2의 적용 과정을 기술하고, 분석 대상 1,999,095통행 기록으로부터 추정된 하차 정류장을 활용하여 추정률과 일치도를 분석한다. 방대한 자료(Big-Data)와 분석 과정의 복잡성으로 인해 Visual Basic for EXCEL과 Oracle SQL Developer(version 4.03)을 활용하였다.

1. 통행사슬 모형 알고리즘 적용

<그림 5-1>과 <그림 5-2>에 본 연구의 통행사슬 모형 적용을 위한 분석 자료 정리를 포함한 알고리즘이 통행사슬 Rule 1과 통행사슬 Rule 2로 구분되어 도식화되어 있다. 알고리즘 적용 과정을 요약하면 아래와 같다.

첫째, 분석 대상 자료인 총 1,999,095승차 통행 기록을 교통카드 번호와 거래 시간순으로 정렬하여, 교통카드 번호별 하루 통행 순서로 자료를 정리한다. 또한, 모든 통합 정류장으로부터 버퍼 거리 400m와 800m 이내에 존재하는 정류장 리스트를 버퍼 거리별로 거리를 포함하여 작성한다.

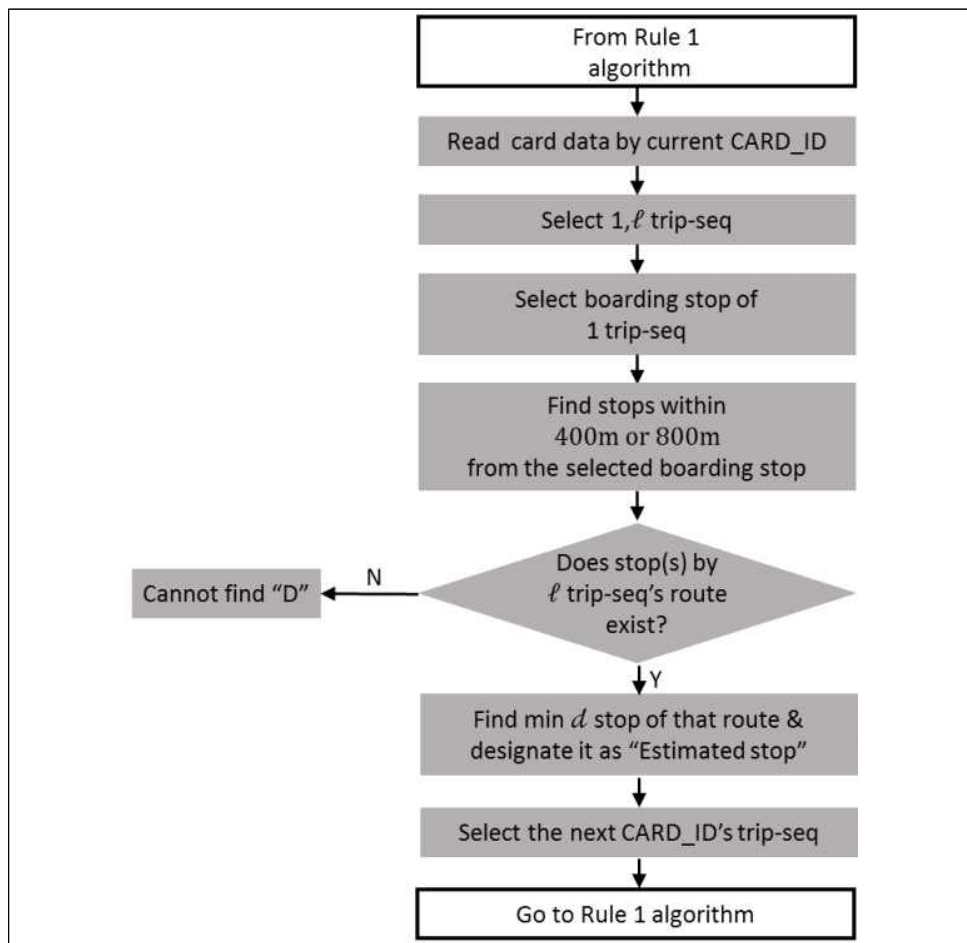
둘째, 통행사슬 Rule 1 알고리즘을 시작한다. 최초 선택된 교통카드 번호 통행 순서 기록 중 두 번째 통행의 승차 정류장으로부터 버퍼 거리 400m와 800m 이내에 존재하는 정류장을 정류장 리스트로부터 도출한 후, 도출된 정류장 중 첫 번째 통행 노선이 통과하는 경우 최단 거리의 정류장을 첫 번째 통행의 하차 정류장으로 간주하고, 리스트 상에 정류장이 존재하지 않는 경우 추정이 불가능한 것으로 판단한다. 이 과정을 선택된 교통카드 번호의 최종 통행까지 반복한 후 선택 교통카드 번호의 최종 통행에서 Rule 1 알고리즘을 종료하고 Rule 2 알고리즘을 시작한다.



〈그림 5-1〉 통행사슬 모형 Rule 1 알고리즘

셋째, Rule 2 알고리즘은 선택 교통카드 번호의 최초 통행과 최종 통행만을 고려한다. Rule 1과 동일하게 최초 통행 승차 정류장으로부터 버퍼 거리 400m와 800m 이내에 존재하는 정류장을 정류장 리스트로 도출한 후, 최종 통행 노선이 통과하는 정류장이 리스트에 존재하는 경우 최단

거리의 정류장을 최종 통행의 하차 정류장으로 결정하고, 리스트 상에 정류장이 존재하지 않는 경우 추정이 불가능한 것으로 판단한다. 최종 통행에 대한 Rule 2 알고리즘이 종료되면 다음 교통카드 번호의 순서화된 통행 중 첫 번째 통행을 선택한 후 다시 Rule 1 알고리즘 과정을 반복 시행한다.



〈그림 5-2〉 통행사슬 모형 Rule 2 알고리즘

본 연구에서 적용된 통행사슬 모형 가정과 알고리즘의 기존 연구와의 차이점은 아래와 같이 요약할 수 있다.

첫째, 목적통행(Linked Trip)의 하차지점이 아닌 수단통행(Unlinked

Trip)의 하차지점을 추정하였다. 따라서 통행의 환승 통행 여부에 대한 분석은 불필요하였다.

둘째, 선(先) 통행/후(後) 통행 혹은 최초 통행/최종 통행 노선과 승차 정류장 거리만을 기준으로 추정이 이루어졌으며, 통행 간의 시간적 차이는 고려 대상에서 제외하였다.

셋째, 선(先) 통행 혹은 최종 통행의 승차 정류장 간 임계 거리 내 존재 여부에 따라 하차 정류장의 추정 가능과 불가능을 결정하였다.

2. 알고리즘 적용 결과 분석

2.1 추정률 결과 분석

제 3 장의 추정률 식 (5)에 의해 정류장 버퍼 거리(임계 거리) 400m와 800m 적용 시 추정률 분석 결과가 <표 5-1>에 제시되어 있다.

<표 5-1> 추정률 분석 결과

(단위 : 통행, %)

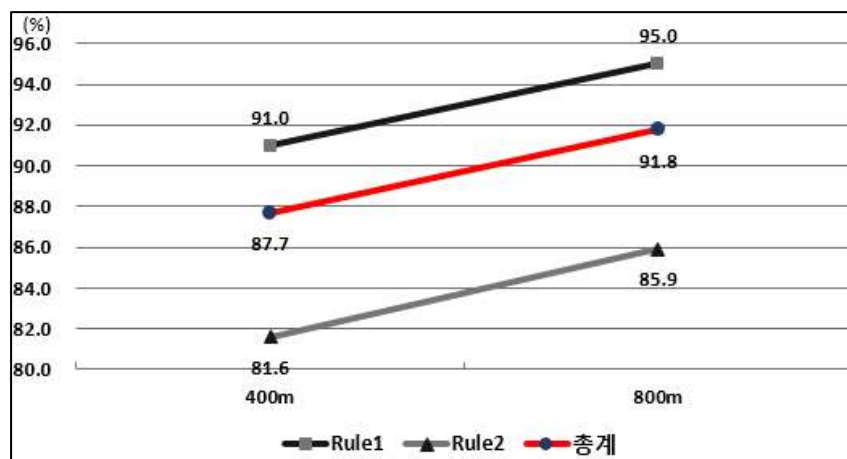
항목			Rule 1	Rule 2	합계
승차		a	1,299,074	700,021	1,999,095
추정 하차	400m	b	1,182,007	570,931	1,752,938
	800m		1,234,217	601,464	1,835,681
추정률 (%)	400m	b/a	91.0	81.6	87.7
	800m		95.0	85.9	91.8

추정률 분석 결과를 요약하면 아래와 같다.

첫째, 추정 대상인 총 수단통행 1,999,095통행 중 Rule 1에 적용 대상인 통행은 1,299,074통행이며, Rule 2 적용 대상 통행(최종 승차 통행)은 700,021통행으로 파악되었다.

둘째, 버퍼 거리에 따라 추정률을 구분하면, 버퍼 거리가 400m인 경우 Rule 1 추정률은 91.0%로 Rule 2를 81.6%로 파악되었고, 버퍼 거리가 2배로 확대된 800m인 경우 Rule 1과 Rule 2의 추정률은 각각 95.0%와 85.9%로 나타나, 400m에 비해 크게 향상되는 것으로 나타났다(<그림 5-3> 참조). 추정률이 향상된 것은 정류장 버퍼 거리의 증가에 따라 버퍼 거리 내 선택 대상 정류장 개수가 늘어난 단순한 사유로 설명할 수 있다. Alsger et al.(2016) 등의 기존 연구와 같이 무한대 버퍼 거리를 가정할 경우 추정률은 100%가 된다. 언급하였듯이 본 연구에서는 비현실적인 무한대 버퍼 거리보다는 버퍼 거리를 벗어난 경우 추정 불가능한 통행으로 가정하였다.

셋째, 버퍼 거리와 상관없이 Rule 1 추정률이 Rule 2의 추정률에 비해 높은 것으로 나타났다. Rule 2는 각 이용자의 마지막 통행을 대상으로 하며, Rule 1은 마지막 통행을 제외한 통행을 대상으로 한다. 다시 말해, Rule 2 대상 통행은 귀가, 퇴근, 하교 등의 통행 유형이 해당되기 때문에, Rule 1의 주요 대상인 출근, 등교, 쇼핑, 비즈니스 등과 같은 유형과 달리 하차를 포함한 통행패턴의 예측이 용이하지 않다. 따라서 Rule 2의 낮은 추정률도 통행 유형의 차이에 비롯된 것으로 판단된다.



<그림 5-3> 버퍼 거리 및 Rule에 따른 추정률

전체적으로 하차 정류장 정보가 부재한 교통카드 자료의 승차 통행에 본 연구의 알고리즘 적용 시 90%에 근접한 하차 정류장 추정이 가능한 것으로 파악된다. 본 연구와 동일한 통행사슬 및 버퍼 거리 가정을 적용한 연구를 살펴보면, Trepanier et al.(2006)가 캐나다 Gatineau의 시내버스 교통카드 자료를 활용하여 추정 불가능한 단일 통행 및 에러 자료를 제외하고 본 연구와 유사한 수치인 추정률 90.5%로 보고하고 있다. Chapleau and Chu(2007)는 버퍼 거리 300m 가정 시 81%, 1km 가정 시 92%로 각각 추정률을 제시하였다. Li(2011)는 중국 Jinan 시의 115번 버스 이용자의 교통카드 자료로부터 본 연구와 유사한 통행사슬 가정을 적용하여 하차 정류장을 추정한 결과 70%(전일) 및 80%(첨두) 추정률을 제시하였다. Munizaga et al.(2010)은 2010년 칠레 Transantiago 교통카드 자료를 활용하여 단일 통행과 오류 자료를 제외하고 추정률 89.4%로 보고하였으며, 다음 해 2009년 3월 및 2010년 6월 자료를 활용하여 연구한 결과 추정률을 91.7%와 91.6%로 각각 제시하였다(Munizaga et al.(2012)). 즉, 타 연구의 결과도 본 연구와 유사한 것으로 판단된다.

〈표 5-2〉 추정률 관련 연구

No.	저자	보정 추정률	No.	저자	보정 추정률
1	Trepanier et al.(2006)	90.5%	4	Li et al.(2011)	70%(전일)
					80%(첨두)
2	Chapleau and Chu(2007)	81%(300m)	5	Munizaga et al.(2012)	91.7% (2009년 3월)
		92%(1km)			91.6% (2010년 6월)
3	Munizaga et al.(2010)	89.4%			-

2.2 일치도 결과 분석

제 3 장의 일치도 식 (9)에 의해 추정률 분석과 마찬가지로 정류장 버퍼 거리(임계 거리) 400m와 800m를 가정하여 분석을 시행하였고, 결과가 <표 5-3>에 제시되어 있다.

<표 5-3> 일치도 분석 결과

(단위 : 통행, %)

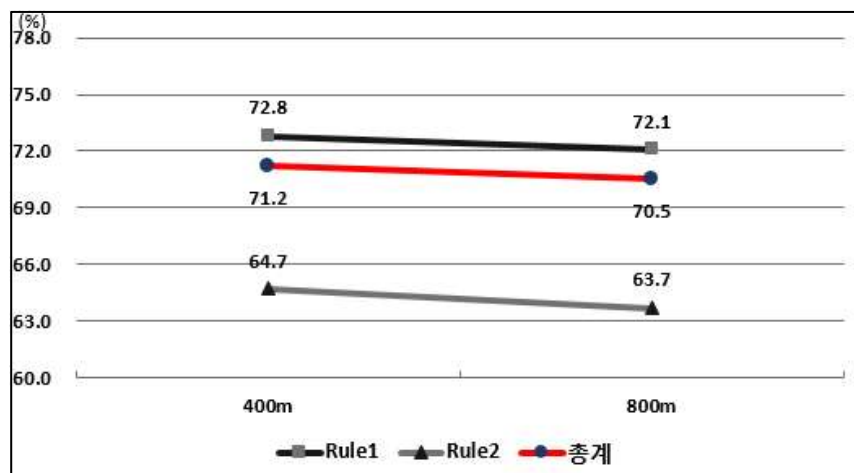
항목			Rule 1	Rule 2	합계
추정 하차	400m	b	766,744	190,660	957,404
	800m		797,310	193,845	991,155
실제 하차=추정 하차	400m	c	558,087	123,322	681,409
	800m		575,008	123,572	698,580
신뢰도 (ρ)	400m	c/b	72.8	64.7	71.2
	800m		72.1	63.7	70.5

일치도는 추정 하차 정류장 중 실제 하차 정류장 정보가 존재하는 통행을 대상으로 분석하였기 때문에 적용된 버퍼 거리에 따라 대상 통행 수에 다소의 차이가 난다. 버퍼 거리 400m 가정 시 실제 하차 정류장 정보가 존재하는 통행 1,044,244통행 중 하차 정류장이 추정된 통행이 957,404통행이며, 800m 가정 시 991,155통행으로 파악되었다. 일치도 계산 결과를 요약하면 아래와 같다.

첫째, 버퍼 거리 400m인 경우 Rule 1 일치도는 72.8%, Rule 2는 64.7%로 파악되었고, 800m인 경우 Rule 1과 Rule 2의 일치도는 각각 72.1%와 63.7%로 400m인 경우에 비해 낮아진 것으로 분석되었다(<그림 5-4> 참조). 버퍼 거리 확대에 따라 추정률이 높아져 실제와 추정이 일치된 통행수가 늘어남에도 불구하고 일치도가 낮아지는 사유는 버퍼 거리의 확대에 의해 추가로 하차 정류장이 추정된 통행이 상대적으로 더 많

아지기 때문이다.

둘째, Rule 1과 Rule 2의 일치도를 비교할 시 상당한 차이를 보이고 있다. 그 사유는 추정률의 차이와 동일한 사유인 최종 통행패턴의 가변성에서 비롯되는 것으로 판단된다.



〈그림 5-4〉 버퍼 거리 및 Rule에 따른 일치도

본 연구의 일치도는 약 70% 정도로 파악되었으며, Rule 1과 Rule 2 모두 버퍼 거리가 확대됨에 따라 낮아지는 것으로 나타났다. 선행연구에서 제시된 일치도를 살펴보면 아래와 같다.

Barry et al.(2002)는 실제 하차 정보가 부재한 뉴욕 지하철 교통카드 자료로부터 하차 정류장의 일치도 분석을 시행하였다. 이를 위해 설문에 의한 통행실태조사(Travel Diary)로부터 도출된 추정 정보와 실제 하차 정보를 비교하여 추정 하차 정보의 일치도가 90%로 제시하였다. Reddy et al.(2008) 역시 뉴욕 지하철과 버스 교통카드 자료로부터 추정된 하차 정보를 통행실태조사 자료상의 하차 정보와 90% 일치도를 보인다고 주장하였다. Nassir et al.(2011)는 Minneapolis-Saint Paul의 AFC 자료로부터 추정된 하차 정보와 APC-VL 생성 자료의 하차 정보를 비교하여 95.4% 일치도를 보였다고 주장하였다. Zhang et al.(2015)은 102명에게

무료 배포된 교통카드이용자의 실제 하차 정보와 추정 하차 정보를 비교하여 75% 일치도를 제시하였다.

실제 하차 정보가 부재한 약점 보완을 위해 실태조사 등과 같은 표본 자료를 활용한 위 연구와 달리 He et al.(2015)와 Alsger et al.(2016)는 실제 하차 정보가 제공된 자료를 활용하여 통행사슬 모형과 가정의 신뢰성 검증을 시도하여 그 결과물로 일치도를 제시하고 있다.

He(2015)는 실제 정류장과 추정 정류장 간 거리가 0m, 즉, 완전히 일치하는 하는 경우 65.8%이며, 거리가 증가할수록 일치도가 증가하여 400m 이내인 경우 79%, 1,000m로 확대할 경우 85%에 도달하는 것으로 제시한다. Alsger(2016)는 800m 버퍼 거리 가정 하에 제시된 존 수준 (Zone-level) 하차지점 추정 시 목적통행 대상 일치도 비율 72.6%를 제시하였고, 정류장 수준(Stop-level) 하차 추정을 위해 실제 정류장과 추정 정류장의 거리가 400m 이내인 경우를 모두 일치한다고 가정하여 약 76%의 일치도를, 800m 이내인 경우 84%의 일치도를 보이는 것으로 분석하였다. 본 연구의 결과를 동일한 가정하에 분석할 경우 400m 적용 시 82%, 800m 적용 시 일치도가 92%로 상승하는 것으로 나타나 보다 양호한 수치를 보이는 것으로 분석되었다. 이는 분석 자료의 속성과 통행패턴의 차이에서 발생하는 것으로 판단된다.

〈표 5-4〉 일치도 관련 연구

No.	저자	보정 일치도	No.	저자	보정 일치도
1	Barry et al.(2002)	90%	4	Zhang et al.(2015)	75%
2	Reddy et al.(2008)	90%	5	He et al.(2015)	65.8% (0m)
					79% (400m)
					85% (1000m)
3	Nassir et al.(2011)	95.4%	6	Alsger et al.(2016)	76% (400m)
					84% (1,000m)

본 연구와 기존 연구에서 적용된 통행사슬 가정의 차이와 분석 대상 교통카드 자료의 속성 및 질적인 차이로 인해 추정률과 일치도의 직접적인 비교는 무리가 있으나, 본 연구에서 제시한 추정률 및 일치도 수치는 기존 연구와 큰 차이가 없음을 알 수 있다.

결론적으로 기존 대중교통계획의 기종점 정보가 대부분 10% 미만의 표본자료로부터 추정되어 참값의 비율이 매우 낮은 것으로 판단되나, 본 연구와 동일한 가정하에 통행사슬 모형을 적용할 경우 90% 이상 하차 정보 추정이 가능하고, 그중 최소 70% 이상 하차 정보가 참값을 나타내는 것을 감안하면, 전체 통행 중 약 60%가 참값인 점은 기존 현장조사 자료 중심의 방법보다는 통행사슬 모형이 신뢰성 관점에서 매우 양호한 방법이라 판단된다.

제 6 장 결 론

본 연구는 하차 정보가 부재한 대중교통이용자의 교통카드 자료로부터 하차지점 추정을 위해 광범위하게 적용되어 온 통행사슬(Trip-chain) 모형의 신뢰성 검증을 목적으로 총 승차 통행 대비 하차 정보가 추정된 통행의 비율인 추정률과 실제 하차 정보가 알려진 통행 중 실제와 추정 하차 정보가 일치되는 비율인 일치도를 분석하였다. 추정률은 통행사슬 모형의 추정 능력을 나타내며, 일치도는 추정된 하차 정보의 정확성을 표현하기 때문에 추정률과 일치도를 파악함으로써 통행사슬 모형의 신뢰성 검증이 가능하다.

추정률과 일치도 분석을 통한 신뢰성 검증을 위해 본 연구는 통행사슬 모형의 2가지 대표적 가정에 대한 수리적인 모형을 구축하였고, 가공된 부산시 교통카드 자료를 대상으로 구축된 모형을 적용하여 하차 정보를 도출한 후 추정률과 일치도 분석을 시도하였다. 부산시 교통카드 자료는 수단 간 무료 환승 제도로 인해 총 승차 통행의 약 30%(부산광역시, 2017) 통행이 실제 하차 정류장 정보가 생성되며, 이로 인해 실제 하차 정류장 정보가 존재하는 통행을 대상으로 일치도 분석이 가능하다. 마지막으로 하차 정보는 Lee et al.(2012)의 SAM을 준용한 정류장 통합 과정을 거쳐 도출된 통합 정류장(Stop-level)을 기초로 수단(Unlinked)통행에 초점을 두어 모든 승차 통행의 하차 정류장을 도출하였다.

적용된 통행사슬 모형의 두 가지 가정 중 첫 번째는 선(先) 통행의 하차지점 추정을 위한 가정이며, 두 번째는 하루 최종 통행의 하차지점 추정을 위한 가정이다. 본 연구에서는 2가지 가정을 각각 통행사슬 Rule 1과 통행사슬 Rule 2로 명명하여 분석을 진행하였으며, 모형식의 버퍼 거리를 도보 1차 및 2차 권역인 400m와 800m로 가정하여 분석을 진행하였다.

본 연구의 통행사슬 모형 적용에 따른 추정률과 일치도 분석 결과를

요약하면 아래와 같다.

우선 추정률 분석을 위한 통행은 총 1,999,095수단(승차)통행으로 파악되었으며, Rule 1 적용 대상은 1,299,074통행, Rule 2 적용 대상은 700,021통행으로 파악되었다. Rule과 버퍼 거리 및 전체 통행의 분석 결과를 요약하면 아래와 같다.

- 버퍼 거리 400m인 경우 Rule 1 추정률은 91.0%, Rule 2는 81.6%로 파악되었고, 전체 추정률은 87.7%로 분석됨.
- 버퍼 거리가 2배로 확대된 800m인 경우 Rule 1과 Rule 2의 추정률은 각각 95.0%와 85.9%로 나타났고, 전체 추정률은 91.8%로 분석됨.
- 버퍼 거리가 증가함에 따라 추정률도 증가하며, 버퍼 거리를 비현실적인 무한대로 할 경우 추정률은 100%, 즉, 모든 통행의 하차 정보가 추정될 수 있음.
- Rule 2에 비해 Rule 1에 해당되는 통행에 대한 추정률이 높은 것으로 파악되었으며, Rule 1과 Rule 2에 해당되는 통행 유형의 차이에서 비롯된 것으로 판단됨.

통행사슬 모형의 추정 능력을 나타내는 추정률이 90%에 근접한 점을 감안하면 정보 추출을 위한 통행사슬 모형 적용은 신뢰할 수 있는 것으로 판단된다.

다음으로 일치도 분석 결과를 살펴보면, 일치도 분석 대상 통행은 버퍼거리 400m 가정 시 하차 정류장이 추정된 1,044,244통행 중 실제 하차 정류장이 알려진 통행이 957,404통행이며, 800m 가정 시 991,155통행으로 감소한다. 일치도 계산 결과를 요약하면 아래와 같다.

- 버퍼 거리 400m인 경우 Rule 1 일치도는 72.8%, Rule 2는 64.7%에 파악되었고, 전체 일치도는 71.2%로 분석됨.
- 버퍼 거리가 2배로 확대된 800m인 경우 Rule 1과 Rule 2의 일치도는 각각 72.1%와 63.7%로 나타났고, 전체 일치도는 70.5%로 분석됨.
- 버퍼 거리 확대됨에 따라 추정률이 높아져 실제와 추정이 일치된 통행수가 늘어남에도 불구하고 일치도가 낮아지는 사유는 버퍼 거리의

확대로 인해 추가로 하차 정류장이 추정된 통행이 상대적으로 더 많아지기 때문인 것으로 판단됨.

- Rule 2에 비해 Rule 1에 해당되는 통행에 대한 일치도가 높은 것으로 파악되었으며, 이 역시 Rule 1과 Rule 2에 해당되는 통행 유형의 차이에서 비롯된 것으로 판단됨.

통행사슬 모형에 의해 추정된 하차 정보가 실제 정보와 동일한 통행이 70% 이상임을 감안하면 매우 제한된 표본으로부터 하차 정보를 추정하는 기존 실태조사에 비해 통행사슬 모형 적용으로 획득할 수 있는 정보의 정확성은 월등한 것으로 분석된다.. 즉, 일치도는 추정된 하차 정보의 정확성을 표현하기 때문에 추정률과 일치도를 파악함으로써 통행사슬 모형은 신뢰할 수 있는 방법론으로 판단된다.

적용된 Rule 1 및 Rule 2의 버퍼 거리 가정에 따라 추정률과 일치도가 변화하나 표본 통행을 대상으로 현장 혹은 설문조사에 의해 파악된 승하차 정보를 기초로 도출된 대중교통이용자의 통행 정보에 비해 교통카드 자료에 통행사슬 모형의 적용을 통해 매우 현실적인 정보 획득이 가능함을 알 수 있으며, 이에 따라 통행사슬 모형의 신뢰성을 확인할 수 있다고 할 수 있다.

본 연구는 분석의 편의를 위해 특정일의 자료를 활용하여 수단통행에 제한하여 통행사슬 모형의 대표적 2가지 가정만을 적용하였으며, 또한, 일부 실제 하차 정보 자료만을 활용한 단점이 있으나, 분석 결과인 추정 하차 정류장 정보에 의해 도출된 존(Zone), 행정동 혹은 행정구별 O-D 도출 및 이를 활용한 분석과 국가교통DB 제공 대중교통 O-D 간의 비교 검증도 가능하다. 나아가서 시간대별 뿐 아니라 분석 대상 자료를 확대할 경우 요일별, 월별 및 연간 대중교통이용자의 통행패턴 분석도 가능하고, 또한, 교통카드 이용자별 통행을 추적하여 개인별 패턴 분석도 가능할 것이다. 궁극적으로 본 연구의 결과가 대중교통 노선 운영 및 계획에 효율적으로 활용될 수 있을 것으로 기대한다.

향후 보다 다양하고 확대된 통행사슬 모형에 대한 가정을 적용하고,

버퍼 거리뿐 아니라 통행 간의 시간 차이를 포함할 경우 보다 다양한 상황에서 추정률과 일치도 분석이 가능할 것으로 판단된다. 또한, 목적통행에 대한 추정과 특정일이 아닌 주간 혹은 월간 자료를 활용할 경우 보다 정확한 하차 정보의 획득이 가능할 것으로 기대하여 본다.

■ 참고문헌

1. 국내문헌

- 김순관, 교통카드 데이터를 활용한 OD 추정 및 활용, 서울시정개발연구원, 연구보고서, 2007.
- 신강원, 통행사슬 구조를 이용한 교통카드 이용자의 대중교통 통행중점 추정, 대한교통학회지, 제34권 제5호, pp. 437-448(2016).
- 조아, 이경희, 조완섭, LDA 기법을 이용한 버스 승객의 잠재적 이동패턴 분석, 한국데이터정보과학회지, 제26권 제5호, pp. 1061-1069(2015).
- 부산광역시, 2016 승객통행량 조사결과, 2017.

2. 해외문헌

- Alsger, A., Mesbah, M., Ferreira, L., Safi, H., "Use of smart card fare data to estimate public transport origin - destination matrix", Transportation Research Record: Journal of the Transportation Research Board No. 2535, pp. 88 - 96(2015).
- Alsger, A., Assemi, B., Mesbah, M., Ferreira, L., "Validating and improving public transport origin - destination estimation algorithm using smart card fare data", Transportation Research Part C: Emerg. Technol. 68, pp. 490-506(2016).
- Bagchi, M., White, P., "What role for smart-card data from bus systems?", Municipal Engineer", 157, pp. 39 - 46(2004).
- Barry, J., Newhouser, R., Rahbee, A., Sayeda, S., "Origin and destination estimation in New York City with automated fare system data", Transportation Research Record: Journal

- of the Transportation Research Board No. 1817, pp. 183 - 187(2002).
- Barry, J., Freiner, R., Slavin, H., "Use of entry-only automatic fare collection data to estimate linked transit trips in New York City", Transportation Research Record, 2112, pp. 53 - 61.(2009).
- Chapleau, R., Trépanier, M., Chu, K.K., "The Validity of Trip-Chain Method Using Smart Card Data", 8th International Conference on Survey Methods in Transport: Harmonisation and Data Comparability, pp. 1-26(2008).
- Chow, W., "Evaluating Online Surveys for Public Transit Agencies using a Prompted Recall Approach", Master's Dissertation. Massachusetts Institute of Technology, pp. 75-90(2014).
- Chu, K.K and R. Chapleau., "Enriching Archived Smart Card Transaction Data for Transit Demand Modeling", Transportation Research Record, 2063, pp. 63-72(2008).
- Cui, A., "Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems", M.S. thesis, MIT, pp. 60-64(2006).
- Devillaine, F., Munizaga, M.A., Trépanier, M., "Detection activities of public transport users by analyzing smart card data", Transportation Research Record: Journal of the Transportation Research Board No. 2276, pp. 48 - 55(2012).
- He, L., Nassir, N., Trépanier, M., Hickman, M., "Validating and Calibrating a Destination Estimation Algorithm for Public Transport Smart Card Fare Collection Systems", No. CIRRELT-2015-52, pp. 1-11(2015).
- Hofmann, M., O'Mahony, M., "Transfer journey identification and

- analyses from electronic fare collection data”, In: Intelligent Transportation Systems, Proceedings IEEE, pp. 34 - 39(2005).
- Jung, J. Y., Sohn, K. m., “Deep Learning Architectrue Used to Forecast Destinations of Bus Passengers from Entry-only Smart-card Data”, IET I.T.S., pp. 1-16(2017).
- Kieu, L., Bhaskar, A., Chung, E., “Mining temporal and spatial travel regularity for transit planning”, Australasian Transport Research Forum (ATRF), 36th, Brisbane, Queensland, Australia, pp. 1-12(2013).
- Lee, S.G., Hickman, M.D., Tong, D., “Stop Aggregation Model Development and Application”, Transportation Research Record, 2276, pp. 38-47(2012).
- Li, D., Lin, Y., Zhal, X., Song, H., Zou, N., “Estimating a transit Passenger Trip Origin-Destination Matrix Using Automatic Fare Collection System”, in Database Systems for Advanced Applications, J. Xu, G. Yu, S. Zhou, and R. Unland, Eds. Berlin, Germany: Springer-Verlag, pp. 502 - 513(2011).
- Lianfu, Z., Shuzhi, Z., Yonggang, Z., Ziyin, Z., “Study on the method of constructing bus stops OD matrix based on IC card data”, Wireless Communications, Networking and Mobile Computing WiCom 2007, pp. 3147-3150(2007).
- Ma, X., Wu, Y.J., Chen, F., Liu, J., “Mining smart card data for transit riders’ travel patterns”, Transport. Res. Part C: Emerg. Technol. 36, pp. 1 - 12(2013).
- McGuckin, N. and Nakamoto Y., “Trips, Chains, and Tours-Using an Operational Definition”, The NHTS Conference, pp. 1-5(2004).
- Munizaga, M.A., Palma, C., Mora, P., “Public transport O - D matrix

- estimation from smart card payment system data”, In: 12th World Conference on Transport Research, Lisbon, 2988, pp. 1-16(2010).
- Munizaga, M.A., Palma, C., “Estimation of a disaggregate multimodal public transport Origin - Destination matrix from passive smart card data from Santiago, Chile.”, Transport. Res. Part C: Emerg. Technol. 24, pp. 9 - 18(2012).
- Munizaga, M.A., Devillaine, F., Navarrete, C., Silva, D., “Validating travel behaviour estimated from smartcard data”, Transport. Res. Part C: Emerg. Technol. 44, pp. 70 - 79(2014).
- Nassir, N., Khani, A., Lee, S.G., Noh, H., Hickman, M., “Transit stop-level origin - destination estimation through use of transit schedule and automated data collection system”, Transportation Research Record: Journal of the Transportation Research Board No. 2263, pp. 140 - 150(2011).
- Nunes, A. A., Dias, T. G., Cunha, J. F., “Passenger Journey Destination Estimation From Automated Fare Collection System Data Using Spatial Validation”, IEEE Transactions on Intelligent Transportation Systems, 17, 1, pp. 133-142(2015).
- Reddy, A., Lu, A., Kumar, S., Bashmakov, V., Rudenko, S., Entry-Only Automated Fare-Collection System Data Used to Infer Ridership, Rider Destinations, Unlinked Trips, and Passenger Miles, In Transportation Research Record: Journal of the Transportation Research Board, No. 2110, Transportation Research Board of the National Academies, Washington, D.C., pp. 128 - 136(2009).
- Trépanier, M., and Chapleau, R., “Destination estimation from public transport smartcard data”, The 12th IFAC Symposium on

Information Control Problems in Manufacturing (INCOM),
Saint-Étienne, France (2006).

Trépanier, M., "Individual Trip Destination Estimation in a Transit
Smart Card Automated Fare Collection System", Journal of
Intelligent Transportation System, 11, 1, pp. 1-14(2007).

Wang, W., Attanucci, J.P., Wilson, N.H.M., "Bus passenger
origin-destination estimation and related analyses using
Automated Data Collection Systems", J. Public Transport. 14,
4, pp. 131 - 150(2011).

Zhang, F., Yuan, N. J., Wang, Y., Xie, X., "Reconstructing individual
mobility from smart card transactions: a collaborative space
alignment approach, Knowledge and Information Systems",
44, 2, pp. 299-323(2015).

Zhao, J., Rahbee, A., Wilson, N., "Estimating a rail passenger trip
origin - destination matrix using automatic data collection
systems", Computer-Aided Civil and Infrastructure
Engineering, 22, pp. 376 - 387(2007).

3. 참고 사이트

부산광역시 버스운송사업조합 : <http://www.busanbus.or.kr/>

네이버 사전 : <http://dic.naver.com>

Analysis of Reliability of Trip-chain Methodology

- Estimation Ratio and Accuracy of Inferred Stop,

Using Busan Smart Card Data -

Su-Jeong, Kim

Dept. of Urban Engineering, Graduate School, Dong-Eui University

Abstract

This study intends to investigate the reliability of trip-chain method, which is widely applied as a method of inferring the departure information from the smart transit card data when there is no alighting information. In order to achieve the research purpose, we first develop a model of two representative assumptions of the trip-chain method to infer the stop-level destination information focusing on the linked-trip only. Then using the inferred results, an analysis is performed to verify the reliability of trip-chain method: first, by evaluating its inferring ability and second, by examining its accuracy thru comparison between the inferred destination and the actual destination.

The two assumptions are termed as trip-chain Rule 1 and trip-chain Rule 2, where the rule 1 is the assumption for inferring the alighting stop of the first trip of two successive trips and the

rule 2 is the assumption for inferring the alighting stop of the last trip of the day. The inferring ability is identified by ratio of the number of trips with the inferred alighting stops and the total boarding trips, while the accuracy is judged by matching the actual alighting stops with the inferred alighting stops.

The model developed is applied to Busan traffic card data, which offer 30% of boarding trips with known actual stop generated by its free transfer system. To derive the stop-level estimation, the stopping integration model (SAM) suggested by Lee et al. (2012) is employed to aggregate all of Busan transit stops suited for the model.

Assuming 400m and 800m buffer distances for each aggregated stops, i.e. walking distance between the aggregated stops, the results of this study can be summarized as follows.

First, it is found that for the 400m buffer distance, the inferred ratios for Rule 1 and Rule 2 are 91.0% and 81.6%, respectively (total 87.7%), while for the 800m, these ratios increase to 95.0% and 85.9% (total 91.8%). Rule 1 assumption clearly shows a better inferring ratio than Rule 2. And as the buffer distance increases, the trip-chain method can infer more destination information. On the whole, one should note that the trip-chain method with the applied assumptions is able to infer close to 90% of the alighting stops of the target boarding trips.

Second, as for the accuracy ratio, it is found that for the 400m, Rule 1 and Rule 2 show 72.8% and 64.7% respectively (total 71.2%), while for the 800m, they become 72.1% and 63.7% (total

70.5%). Contrary to the inferring ability, as the buffer distance increase, the method generates more inaccurate information, while Rule 1 produces more accurate information than Rule 2 as above inferring ability. The result nevertheless confirms that the trip-chain method can generate much more true information than the conventional methods which rely mostly on the field or questionnaire survey data.

All in all, this study confirms the reliability of the trip-chain method in terms of its inferring ability and accuracy of the inferred information. It is expected that based on the more diverse assumptions along with the refined trip-chain method, one can improve the inferring and the accuracy ratio. It is also expected that the refined method can consider not just the distance dimension but also the time dimension so as to improve its inferring ability and accuracy.

Keywords : *Public transport system, Trip-chain method, Smart card, inferring Alighting Point, Stop Aggregation, Big-Data Analysis*

■ 감사의 글

정말 많은 분들의 도움이 있었기에 부족한 제가 무사히 석사학위논문을 마칠 수 있게 되었습니다.

먼저 논문이 완성되기까지 연구의 방향에 대하여 아낌없이 조언과 도움주신 **신용은 교수님**께 감사드립니다. 교수님과 나누었던 길고 짧았던 대화가 모두 교통분야에 대한 저의 철학이 되었습니다. 또 때로는 학문과 인생의 선배 입장에서 제게 해주신 따뜻한 말씀들 가슴속에 새기고 앞으로도 포기하지 않고 학문에 있어서도 인생에 있어서도 정진하도록 하겠습니다.

또한 많은 가르침과 도움을 주신 **김홍관 교수님, 김가야 교수님, 윤상복 교수님, 윤항록 교수님, 김형보 교수님, 백태경 교수님**께도 감사의 말씀을 전하고 싶습니다.

후배의 논문에 눈빠지게 봐주시느라 고생하신 **채윤 선배, 민우 선배, 광희 선배**께도 감사드립니다.

대학원 마지막 한해 동안 논문과 놓치는 부분을 챙겨주고 학교에 같이 있었던 유일한 선배 **현주 언니** 고마워요!

석사 동기 **종언이 오빠** 우리 먹느라 고생했다. 아직 살 덜 찐거 같은데 더 분발해서 대차게 살찌자.

내 해결사 술지(**임원희**) 님은 **선미 언니** 1년 동안 내 팔걸이가 없어 불편했다. 같이 학교에 있지 않았지만 연락으로 언제든지 물어보면 도와줘서 고마웠어. 방탄소년단 파이팅!

버스조합 스쟁이 **호영이 언니** 논문쓰는 동안 밥과 술을 사줘서 고맙고 한동안 못 놀았던거 이제 놀 수 있어요!

올해 26살을 맞이한 **엄버딸 강패트, 두두, 팽이보스, 주설, 표세**. 갯돈 모

은겉로 얼른 어디든 놀러가자. 밀린 겹돈 얼른 내보도록 할게. **안수, 찬희**
운동다니기 전에 맛있는거 먹으러가자.

4년 동안 묵묵히 옆에서 응원해주고 이해해준 **지용이** 고마워. 이제 자주
만나고 고기먹으러가자.

마지막으로 저의 전부인 **가족들**. 앞으로도 항상 정직하게 노력하며 자랑
스러운 딸이 될 수 있도록 하겠습니다. 우리 가족 앞으로도 항상 건강하고
서로 의지하며 행복하게 삽시다. 정말 사랑합니다.

학위를 받는 것이 끝이 아니라 새로운 시작이라는 마음으로 매사에 넓은
포용력과 관심, 배려를 아끼지 않는 삶이 되도록 더욱 노력하겠습니다.

감사합니다.

2017년 12월 김수정 拜上

[부록 1] 부산시 마이비 카드 98개 속성 정보 리스트

No.	속성	No.	속성	No.	속성
1	TRX_SERNO	27	FARE_APPLY_DATE	53	HS_DIS_AMT
2	OUT_ORG_CODE	28	DIS_RATE	54	HS_SUM_AMT
3	OUT_ORG_SEQ	29	HS_RATE	55	HS_LANE_TYPE
4	AREA_CODE	30	ST	56	HS_INFO
5	PCARD_CLASS_CD	31	A_EXTRA_ID	57	HS_REFUSE_CODE
6	MERANT_NO	32	A_CNT	58	BEF_TM_NO
7	TRANSP_BIZR_ID	33	A_AMT	59	BEF_HS_CNT
8	CARD_NO	34	A_BASE_AMT	60	BEF_TRX_AMT
9	RID	35	S_EXTRA_ID	61	BEF_ST
10	A_CARD_NO	36	S_CNT	62	BEF_VEHICLE_TYPE
11	ON_OFF_FLAG	37	S_AMT	63	BEF_COURSE_NO
12	TRX_DATE	38	S_BASE_AMT	64	BEF_LANE_NO1
13	TRX_AMT	39	C_EXTRA_ID	65	BEF_LANE_NO2
14	CARD_BAL	40	C_CNT	66	BEF_LANE_TYPE
15	CARD_NT	41	C_AMT	67	BEF_TRX_DATE
16	COURSE_NO	42	C_BASE_AMT	68	BEF_RIDE_AREA_CODE
17	EB_LANE_NO	43	E_CNT	69	CHRG_AFTER_CARD_BAL
18	LANE_TYPE	44	E_AMT	70	CHRG_CARD_NT
19	DRIVE_CNT	45	E_BASE_AMT	71	CHRG_AMT
20	TM_NO	46	TR_TYPE_CD	72	CHRG_SAM_ID
21	TM_NT	47	TR_KEY_VERSION	73	CHRG_SAM_NT
22	DRIVER_ID	48	E_CASH_COM_ID	74	CHRG_TR_TYPE
23	VEHICLE_TYPE	49	CARD_ID	75	CHRG_TRX_DATE
24	EB_VEHICLE_NO	50	SAM_ID	76	BOKGI_FLAG
25	TR_INFO	51	SAM_NT	77	BOKGI_TYPE
26	MULTI_RIDE_CD	52	SAM_TOT_TR_CNT	78	FILE_DATE

<표 계속>

No.	속성	No.	속성	No.	속성
79	CITY_IN_OUT_CD	86	SAM_IND_TR_CNT	93	U1000_FILE_NAME
80	EXTRA_CHARGE_CD	87	SAM_CUM_TR_TOT_AMT	94	E_CODE
81	CARD_TYPE	88	UNIT_SAM_CUM_TR_TOT_CNT	95	SEND_FLAG
82	CARD_KIND	89	SIGN	96	C_ECODE
83	CARD_STATUS	90	RIDE_AREA_CODE	97	C_CLEARDATE
84	STUDENT_EXPIRE	91	HS_FLAG	98	COLLECT_DATETIME
85	FARE_INFO	92	HS_CNT	-	