

버스정보시스템 데이터를 활용한 교통카드 데이터 오류보정 알고리즘

탁화정¹, 송혜인², 손상훈¹

¹ 제주연구원 환경도시연구부

² ㈜배가스



목 차

버스정보시스템 데이터를 활용한
교통카드 데이터 오류보정 알고리즘



배경 및 목적



기존연구 고찰



오류유형화 및
오류보정 알고리즘



버스정보시스템
검증 및 오류파악



결론 및 한계점

배경 및 목적

버스정보시스템 데이터를 활용한 교통카드 데이터 오류보정 알고리즘

- 교통카드 데이터는 대중교통분야에서 활용도가 큼
- 하지만, 교통카드 데이터 수집 특성 상 이용자의 하차태그 행태 또는 물리적, 환경적 요인들로 인해 오류 및 결측 발생
- 신뢰성 높은 데이터 제공을 위해 결측 및 오류에 대한 보정관련 연구는 필연적임
- 기존 연구들의 오류보정 방법론은 통행이력 자료를 통한 역추정 방법 또는 필터링을 통한 오류데이터 삭제
- 통행자의 이력데이터가 없을 경우 불가능, 교통카드 데이터 수 감소



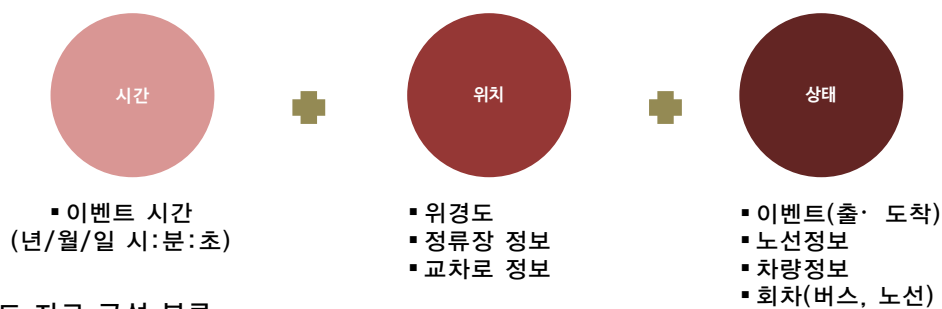
- **버스정보시스템 데이터를 활용한 교통카드 데이터 오류보정 알고리즘 제안하고자 함**
- **제주지역에서 수집된 실측자료(버스 승하차 조사)와 비교하는 방법으로 검증작업 수행하고자 함**

1

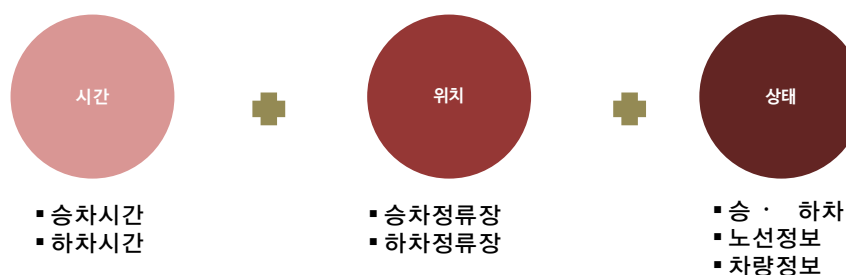
버스정보시스템 및 교통카드 자료

버스정보시스템 데이터를 활용한 교통카드 데이터 오류보정 알고리즘

● 버스정보시스템 자료 구성 분류



● 교통카드 자료 구성 분류



2

기존연구 고찰

버스정보시스템 데이터를 활용한
교통카드 데이터 오류보정 알고리즘

논문명	내용	보정 방법	검증
대중교통 OD구축을 위한 대중교통카드 데이터의 오류와 결측 분석 및 보정에 관한 연구	<ul style="list-style-type: none"> • 통행수요 분석과 관련한 오류나 결측 유형 제시 • 교통카드 결측자료 보정 방안 구축 	<ul style="list-style-type: none"> • 개인별 Trip chain 환승정보 활용 • 개인별 반복 통행 자료 활용 • 일일 통행의 회귀적 특성 활용 • 노선별 통행패턴 활용 • 카드 이외 승차자의 통행량을 보정 • 교통카드 이용비율 활용 	결측자료 제거 후 정류장 OD 구축 → 임의로 10%, 20%, 30% 결측 발생 후 비교
교통카드 데이터를 이용한 버스 승객 대기시간 최소화 알고리즘 개발	<ul style="list-style-type: none"> • 교통카드 데이터를 활용하여 정책적 대기시간 제약, 운행 대수 제약조건, 노선특성을 반영한 승객들의 대기시간 최소화하는 배차간격 조절 알고리즘 개발 • 오류: 기대치 오류와 논리 오류로 구분함 • 결측: 승차결측과 하차결측으로 구분함 	<ul style="list-style-type: none"> • 오류값을 포함한 칼럼을 필터링 기법을 통해 제외함 • 하차결측의 경우 버스 이용자의 연계 통행자료를 이용해 유실된 하차 결측 정보를 역추정함 	검증작업 미수행

한계점

- 통행이력정보 생성시 상당한 시간적·하드웨어적 비용수반
- 기존 보정방법은 개인 통행에 있어 반복되는 패턴이 존재해야 함
- 신규 교통카드의 경우 오류보정 불가능
- 교통카드 원 데이터 삭제

3

교통카드 오류유형

버스정보시스템 데이터를 활용한
교통카드 데이터 오류보정 알고리즘

● 교통카드 5가지 오류유형

구분	오류명	오류정의
정류장 정보 오류	오류 1	승차정류장 위치오류 승차정류장 위치가 실제 버스가 정차한 정류장 위치와 불일치
	오류 2	하차정류장 위치오류 하차정류장 위치가 실제 버스가 정차한 정류장 위치와 불일치
	오류 3	승차정류장 기재오류 승차정류장 ID 중 (주)한국스마트카드사에서 제공한 정류장ID 정보와 불일치
	오류 4	하차정류장 기재오류 하차정류장 ID 중 (주)한국스마트카드사에서 제공한 정류장ID 정보와 불일치
	오류 5	승하차정류장 동일 이상오류 승차정류장과 하차정류장이 동일

- 오류 1, 오류2, 오류5는 본 연구에서 제안한 오류보정 알고리즘을 통해 보정 가능함
- 오류 3, 오류 4는 오류특성에 따라 보정 가능 여부를 파악함

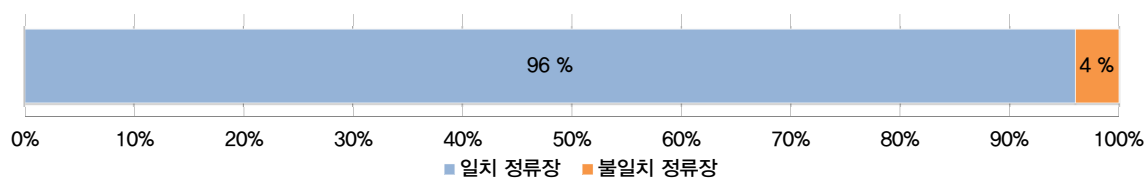
4

버스정보시스템 검증

버스정보시스템 데이터를 활용한
교통카드 데이터 오류보정 알고리즘

● 제주특별자치도 버스정보시스템과 2018년 버스승하차 조사 비교 결과

- 비교 날짜: 18/01/15, 18/01/17, 18/01/18, 18/01/21, 18/01/24, 18/01/26, 18/01/29, 18/01/30, 18/02/01, 18/02/08, 18/02/10, 18/02/11, 18/02/12, 18/02/13, 18/02/19, 18/02/20, 18/02/21, 18/02/22, 18/02/23, 18/02/24, 18/02/25
- 비교 노선 및 회차: 48회(주중 : 33, 주말 : 15, 조사구간: 전체구간, 날씨 : 눈, 비 제외, 노선유형별)
- 비교 정류장: 1,070개 정류장
- 비교 결과



7

교통카드 오류파악 및 보정결과

버스정보시스템 데이터를 활용한
교통카드 데이터 오류보정 알고리즘

● 교통카드 오류파악

- 시간적 범위 : 2018년 1월 ~ 2월
- 공간적 범위 : 전체정류장, 전체노선

구분		오류명	오류개수	보정개수	보정률
정류장 정보 오류	오류 1	승차정류장 위치오류	409,454	409,454	100%
	오류 2	하차정류장 위치오류	161,661	161,661	100%
	오류 3	승차정류장 기재오류	43,605	1,430	3.3%
	오류 4	하차정류장 기재오류	33,546	1,255	3.7%
	오류 5	승하차정류장 동일 이상오류	243,668	114,344	46.9%

8

- 1 제주지역에서 수집된 실측자료를 활용해 버스정보시스템의 정확도를 검증**
- 2 버스정보시스템 데이터를 활용한 교통카드 데이터 오류 보정 알고리즘 제안**
→ 추가적인 데이터 수집 및 통행자료 생성 X, 교통카드 원 자료 유지
- 3 2018년 버스승하차 조사와의 비교 시 표본의 모집단 대표성문제**
→ 버스승하차 조사는 최대한 랜덤하게 조사하고자 하였으나, 조사원의 조사 선호 시간대와 노선이 존재함
- 4 버스정보시스템 데이터 정확도에 따라 오류보정의 정확도가 달라짐**
→ 버스정보시스템 데이터 정확도가 교통카드 데이터보다 더 정확하다는 가정하에 본 연구에서 제안한 알고리즘이 성립됨



2018년 버스승하차 조사 결과를 활용하여 오류보정 알고리즘 검증 수행

(제주특별자치도 2018년 버스승하차 조사 결과와 보정 전후 승차정류장별 승차인원을 비교하는 방식으로 검증 수행 예정)