

제주 택시 텔레매틱스에 기반한 택시 승차지점 분석 기법 †

Analysis scheme for customer pick-up points based on the Jeju Taxi Telematics system

이정훈, 박경린
Junghoon Lee, Gyung-Leen Park
제주대학교 전산통계학과
{jhlee, glpark}@cheju.ac.kr

요약

본 논문은 제주 택시 텔레매틱스 사업의 운영 결과 축적된 히스토리 정보를 기반으로 택시들의 운행기록을 분석하여 승객들이 택시를 찾는 지점을 분석하는 기법을 제시하고 이를 바탕으로 시간별 지역별 링크별 승차 패턴을 분석한다. 이를 위하여 택시의 상태도 천이에서 승차지점을 추출하였으며 해당 데이터를 데이터베이스 테이블에 저장하였다. 이후 승차지점들을 그루핑하여 승차 패턴의 추이를 발견하고 이에 대한 분석을 수행하였다. 이 분석 데이터는 택시들을 지역별 시간대별로 승객이 많이 찾는 위치로 이동시키고 택시의 공차 운행율을 감소시키는 배차 방식을 개발하여 택시의 영업 수익 증대와 승객의 택시 대기시간 감소를 기할 수 있다.

1. 서론

제주 택시 텔레매틱스 시스템은 제주 지역에서 2006년 5월에 시작되었으며[1] 택시의 효율적인 배차를 위하여 택시의 위치를 파악하고 고객의 콜에 대해 가장 빠른 택시를 배차할 수 있도록 한다[2]. 이를 위하여 택시에 텔레매틱스 장치를 부착하였으며 CDMA 통신을 통해 중앙의 관제서버에게 택시의 상태를 보고하도록 한다. 이 시스템은 효율적인 배차, 카드 결제, 서비스의 확대 등 기본적인 목적 이외에도 택시들의 위치에 대한 다양한 로그 데이터를 축적함으로써 고부가가치의 데이터를 생성할 수 있는 프레임워크를 제공할 수 있다.

각 택시의 텔레매틱스 장치들이 매 1분마다 보고하는 레코드는 GPS 수신기에서 입력된 시간, 경도, 위도, 속도, 방향 등의 값과 더불어 택시 텔레매틱스에서 추가한

택시 ID와 현재 승객이 타고 있는지를 나타내는 status 정보들이 1차적으로 포함된다[3]. 이러한 레코드를 받아 맵 매칭 프로그램은 이 좌표값에 해당하는 링크를 검색하여 link 필드에 추가한다[4]. 또 링크 상에서 어느 정도 비율에 있는지를 계산하여 pos 필드에 저장한다. 이렇게 구성된 레코드들로 report 테이블이 구축되었으며 가입 택시의 수에 따라 가변적이기는 하지만 레코드들이 축적되고 있다.

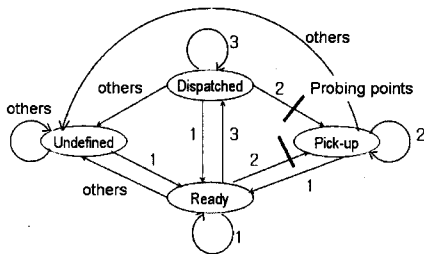
본 논문에서는 이러한 데이터를 기반으로 택시 고객의 탑승 지점을 분석하여 보다 효율적인 택시의 배차 방식을 제안할 수 있는 기본 데이터를 추출하고자 한다. 실제 택시들의 운행기록을 조사하면 80%에 달하는 레코드들이 공차상태, 즉 승객이 없는 상태임을 보이고 있다. 이를 개선하여 공차율을 줄이려면 택시들이 시간대

† 본 연구는 지식경제부 및 정보통신진흥연구원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (IITA-2008-C1090-0040)

별 공간별로 승객들이 많이 차량을 탑승하는 위치로 도달할 수 있도록 유도하여야 한다. 따라서 본 논문은 택시들의 운행 기록 데이터에서 승객들이 탑승 히스토리를 추출하여 데이터베이스화하며 이를 기반으로 공차들의 위치를 지정함으로써 공차율을 줄이고 승객들의 대기시간을 감소시킨다.

2. 택시 승차 지점의 추출

승객이 탑승한 지점을 추출하기 위해 report 테이블 내의 모든 레코드들을 차량 id에 의해 그루핑하고 이를 시간순으로 정렬한다. 이후 승객 탑승 상태로 바뀐 지점과 승객 미탑승 상태로 바뀐 지점을 추출한다. (그림 1)은 각 택시들의 상태 천이를 보이고 있으며 그림에서 보는 바와 같이 승객이 탑승한 Pick-up 상태로 천이한 순간의 시간과 장소를 추출할 수 있다[5]. 이 천이는 택시가 공차였다가 승객을 태우거나 혹은 승객의 콜에 따르는 배차를 받아 승객을 태울 수 있으며 그림에서 Probing points라고 명시된 부분에 해당한다.



1: Empty 2: Passenger on D: Dispatched 4: Rest

(그림 2) 승차지점의 추출

이와 같이 추출된 정보들을 테이블에 저장하기 위해서는 (표1)에서 보인 geton 테이블이 생성되어야 한다.

(표 1) 승차정보 저장 테이블 생성

```

DROP table IF EXISTS geton;
CREATE table geton (
  tstamp DATETIME NOT NULL,
  x      double(11,10),
  y      double(11,10),
  link   int(6) NOT NULL
);
  
```

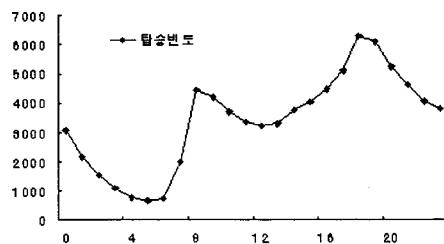
생성된 테이블에 정보를 삽입하기 위해서는 순차적인 정보를 처리하는 프로그램을 작성하여 일차적으로 승차지점에 대한 레코드들을 추출하여 이를 geton 테이블에 INSERT하는 SQL 문으로 변환하여 파일에 저장한다. 이후 system 함수에 의해 또 하나의 쉘 상에서 mysql 을 수행시켜 생성된 SQL 문들이 수행되어 geton 테이블에 새로운 레코드들이 추가 되도록 한다.

이렇게 생성된 테이블을 기반으로 하여 승차지점에 대한 확률분포와 시간대별 탑승 분포를 도시하면 (그림 2)와 같다. (그림 2)는 각 승차지점의 위치를 제주도 도로 네트워크 상에 1/1000 확률로 플로팅한 결과이다. 제주시의 경우 시청등 관공서가 밀집한 구제주 지역, 주택가 및 호텔, 식당들이 모인 신제주 지역, 그리고 공항 지역 등에서 많은 승차가 이루어지고 있다.



(그림 3) 승차지점의 확률분포

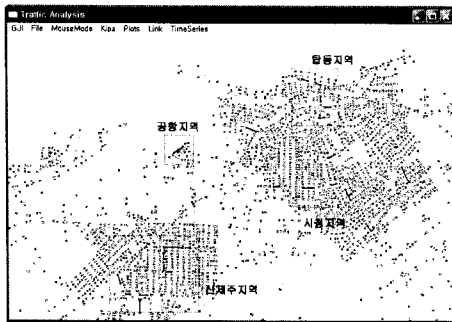
또 (그림 3)은 시간대별로 승차에 대한 빈도를 추출하여 정렬한 결과로서 출근 시간대와 퇴근시간대 또 심야 시간대에 많은 승차가 일어남을 알 수 있다.



(그림 4) 승차 시간대의 분석

3. 승차 패턴에 대한 분석

승차 패턴의 분석은 크게 도로별, 지역별 승차에 대한 시간대별 분포로 구성된다. 택시 승차 가능성이 높은 지역은 유동인구가 많은 지역으로서 제주시에서는 시청, 신제주, 탐동, 공항 지역 등이 선택될 수 있으며 (그림 4)는 이들 지역에 대한 분포를 도시하고 있다.



(그림 5) 지역별 그루핑

3.1 도로별 승차 링크의 분석

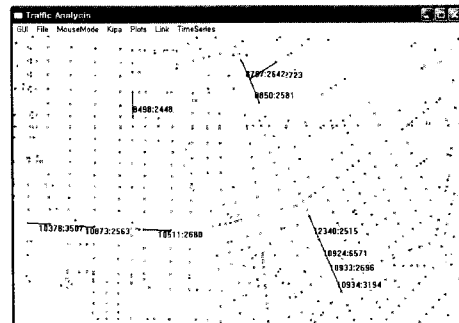
도로별 탑승횟수를 분석함에 있어서 각 도로에는 고유한 링크 번호가 부여되어 있으며 도로의 길이와 같은 특성에 따라 승차 횟수가 영향을 받기도 한다. 링크에 대한 분석은 (표 2)의 SQL 문장에 의해 수행되었다.

(표 2) 링크에 대한 분석

```
drop table if exists frequency;
create table frequency select link, count(link)
as freq from geton group by link;
select * from frequency order by freq desc;
drop table frequency;
```

(그림 5)는 대표적으로 시청 지역의 링크별 탑승회수를 보이고 있다. (그림 5)에서 링크 옆의 숫자는 택시가 승객을 태운 도로의 번호, 순위, 횟수 등을 나타낸다. 각 세부 지역을 확대하여 링크의 아이디와 빈도를 세부적으로 확인할 수 있다. 분석 결과에 따르면 공항지역이 가장 많은 승객을 태운 지점으로 나타난다. 833회의 승객을 태운 15456번 링크의 위치도 공항지역에 포함되어 공항지역에 포함된 5 개의 링크가 모

두 100회 이상의 승객 탑승지점으로 나타나고 있다. 그 다음 탐동 지역이 7554, 7594 번 링크를 포함하고 있는데 이들은 각각 389, 304번의 승객 탑승을 하였다. 신제주 지역의 경우는 연상로 주변과 신흥아파트 단지 주변에서 탑승이 많이 이루어졌음을 알 수 있다. 또 5·16도로 변의 시청과 법원 근처에서 비교적 고른 승객탑승이 일어났음을 보여주고 있다.

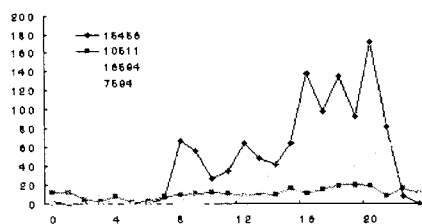


(그림 6) 시청 지역의 예

이와 아울러 각 링크는 시간대에 탑승 횟수에 영향을 받는데 각 지역의 대표적인 도로의 탑승 횟수를 시간대별로 분석하기 위해서는 (표 3)에서 보는 SQL 문장을 사용한다. SQL 문장에서 각 링크들의 ID가 직접적으로 명시되어 있다. 이 문장을 수행한 결과는 (그림 6)에서 보인다.

(표 3) 링크의 시간대별 탑승률 분석

```
select hour(timestamp), count(*) from geton
where link = 15456 group by hour(timestamp);
select hour(timestamp), count(*) from geton
where link = 10511 group by hour(timestamp);
```



(그림 7) 링크의 시간대별 탑승률 추이

위 그림에서 에서 지역별 탑승횟수가 가장 높은 링크는 15456(공항), 7594(중앙로), 16594(신제주), 10511(시청)이다.

3.2 시간대별 승차 링크의 분석

시간대별 분석을 위해 (그림 3)의 그래프를 기준으로 시간대를 7-9시, 9-16시, 16-20시, 20시-7시까지의 4 개 구간으로 설정하였다. 이는 출퇴근 시간과 비행기 착륙시간 및 심야시간 등을 반영할 수 있다. (표 4)는 출근 시간대에 대한 승차율을 분석하는 SQL 질의문을 보이고 있다.

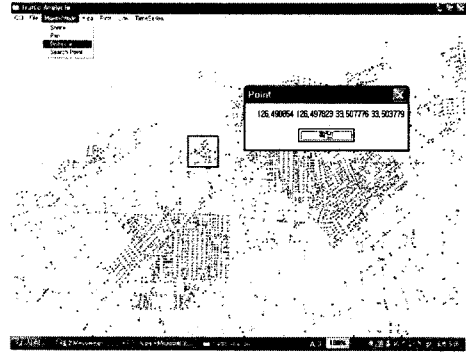
(표 4) 출근시간대의 승차율 분석

```
drop table if exists frequency;
create table frequency select link, count(link)
as freq from geton
where ((hour(timestamp) >= 7) and
(hour(timestamp) < 9)) group by link;
select * from frequency
order by freq desc limit 20;
```

시간대별 분석 결과에 의하면 항공운항이 종료되는 밤 시간을 제외하고는 공항 부근에서 가장 많은 탑승이 발생하며 출근시간에는 구제주 지역이, 퇴근, 저녁 및 심야 시간에는 신제주 지역에서 많은 링크들이 상위에 랭크되는 현상을 보이고 있다.

3.3 주요 지역별 좌표값과 승차 빈도 분석

각 지역의 전체적인 추이를 분석하기 위해서는 링크뿐만 아니라 영역에 대한 분석이 필요하다. 이를 위하여 (그림 7)에서 보는 바와 같이 영역을 설정하여 이 범위 내에서 승차한 횟수를 카운트하였다. 영역은 앞에서와 마찬가지로 공항 지역, 신제주 지역, 탑동 및 중앙로 지역, 시청 지역을 대상으로 하였다. 그림에서 보는 바와 같이 마우스 모드는 놓리는 점과 띄어지는 점의 컴퓨터 모니터 상에서의 좌표를 제주도 지도 내에서의 경위도 좌표로 환산하여 해당 영역의 사각형 틀을 얻어낸다. 이 과정에서 화면 좌표계와 실제 위치 상에서의 좌표 변환이 수행된다.

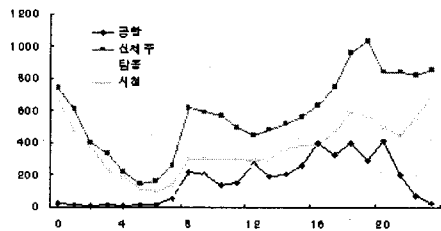


(그림 8) 좌표값과 영역의 추출

이와 같이 추출한 좌표값들에 의해 (표 5)의 SQL 문장을 수행하여 각 지역별 탑승 빈도를 구할 수 있다.

(표 5) 영역별 질의를 위한 SQL 문

```
# 탑동
select hour(timestamp), count(*) from geton where
(x > 126.491218) and (x < 126.511511) and
(y > 33.499965) and (y < 33.506549)
group by hour(timestamp);
# 신제주
select hour(timestamp), count(*) from geton where
(x > 126.478661) and (x < 126.495529) and
(y > 33.483654) and (y < 33.495529)
group by hour(timestamp);
```



(그림 9) 지역별 승차패턴의 분포

(표 5)의 질의를 네 개의 지역에 모두 적용한 후 그 추이를 도시하면 (그림 8)과 같다. 결과에서 보는 바와 같이 전체적인 탑승횟수는 전 시간구간에서 신제주 지역이 가장 많다. 최근 인구의 집중과 기존의 유흥지 근처에서의 꾸준한 탑승이 그 원인으로 분석된다. 공항은 항공기 도착이 본격적

으로 시작되는 8시부터 21시까지 많은 탑승횟수를 보인다.

4 결론

본 논문에서는 최근 대두되고 있는 이동이력 데이터의 처리기법 개발 동향에 따라 [6] 제주 택시 텔레매틱스 시스템에서 생성된 히스토리 데이터를 분석하여 승차지점에 대한 패턴 분석을 수행하였다. 분석 과정은 먼저 상태도에서 탑승에 해당하는 천이부분을 프로빙하여 택시 승차의 시각과 지점 좌표를 추출하였다. 이 정보를 바탕으로 신제주, 시청, 탑동, 공항 영역의 탑승 패턴을 조사하는 SQL 질의를 작성하였으며 그 결과를 플로팅하였다.

분석 결과는 각 지역마다 각각의 특성, 즉 관광관련, 행정관련, 주거 및 서비스 밀집 지역 등을 반영한 패턴을 보이고 있다. 이러한 시간적 공간적 데이터는 향후 택시의 위치를 보다 승객이 많은 지역으로 미리 배치시켜 공차율을 줄이고 승객의 대기시간을 감소시킬 수 있다. 이와 아울러 추후에는 링크의 통행속도를 산출하는 기법에 대한 연구도 진행될 예정이다[7].

참고문헌

- [1] J. Lee, G. Park, H. Kim, Y. Yang, P. Kim, and S. Kim, "A telematics service system based on the Linux cluster," *Lecture Notes in Computer Science*, Vol. 4490, pp. 660–667, 2007.
- [2] Z. Liao, "Real-time taxi dispatching using global positioning systems," *Communication of the ACM*, Vol. 46, pp. 81–83, 2003.
- [3] J. Lee, G. Park, "Design of a connectivity analysis scheme for hybrid telematics networks," accepted at *International Conference on Embedded Systems and Applications*, Las Vegas, 2008.
- [4] S. Brakatsoulas, D. Pfoser, R. Salas, and C. Wenk, "On map-matching vehicle tracking data," *Proc. 31st VLDB Conference*, 2005.

- [5] 이정훈, 권상철, "상태도에 기반한 택시 텔레매틱스 히스토리 분석," 『한국공간정보시스템학회지』, 제10권 1호, 2008년 3월, pp. 41–49.
- [6] J. Krumm, "The geographic context browser," *International Workshop on Exploiting Context Histories in Smart Environments*, 2005.
- [7] S. Lee, B. Lee, and Y. Yang, "Estimation of Link Speed Using Pattern Classification of GPS Probe Car Data," *Proc. International Conference on Computational Science and its Applications*, pp. 495–504, 2006.