

위치이력 데이터를 이용한 공택시 배치에 관한 연구[†]

A study on the placement of empty taxis based on the location history data

이정훈 · 박경린

Junghoon Lee, Gyung-Leen Park

제주대학교 전산통계학과

{jhlee, glpark}@cheju.ac.kr

요 약

본 논문은 택시들의 승차율을 높이고 승객들의 대기시간을 최소화하기 위하여 제주 택시들의 이동이력 데이터를 기반으로 하여 공차들을 승객을 만날 가능성이 많은 지역으로 배치하는 기법을 제시한다. 이동이력 데이터에 포함된 공차 보고와 승차 보고 수 사이의 스케일 차이를 극복하기 위하여 전체수와 영역내 함에 대한 비율로 정규화하는 방법을 설명하고 장단점을 분석한다. 또 시간대별, 요일별, 주간별 택시들의 승객 대기시간에 대한 통계 데이터에 기반하여 가장 수요와 공급이 적정하게 유지되는 시간구간을 발견하고 이 구간에 대한 택시 분포와 현재의 택시 분포의 차이에 의해 수요보다 공급이 많은 곳의 택시를 재배치한다.

1. 서론

현재 제주도에서 운영되고 있는 택시 텔레매틱스 시스템은 택시들의 실시간 위치 파악에 의한 효율적인 배차라는 목적이외에도 택시들의 위치 이력 데이터를 축적하고 있다[1]. 즉, 실시간 위치 파악을 위해 각 택시로부터 1분 혹은 3분마다 현재의 경위도 위치를 보고 받기 때문에 택시들의 이동 내역이 축적되고 있다. 경위도 좌표뿐만 아니라 속도, 방향, 승객의 탑승 여부 등 부가적인 정보들도 축적되고 있기 때문에 이력 데이터는 승객들의 탑승 빈도 패턴, 탑승 거리 패턴, 링크의 통행 속도, 공차율, 배차시의 이동 거리, 배차 후 승객에게 이동할 때까지의 시간 등 다양한 정보들을 추출할 수 있다. 택시들은 공차시에 승객을 찾기 위해 일정한 속도-대부분 주행시보다는 낮은 속도-로 이동하거나 특정 장소에서 거의 정지하여 기다린다. 한 예로 제주국제공항에

는 항공편을 통해 도착한 승객들을 기다리는 빈 택시들이 모이게 된다. 사실, 빈 택시의 경우, 특정속도로 주행하는 것이 승객을 찾는 데 유리한지 아니면 특정한 장소에서 기다리는 게 유리한지에 대해 택시들의 공차율 차이를 기반으로 위치이력 데이터에 대한 분석을 시도한 바 있으나, 이에 대한 구체적인 통계적인 해답을 찾을 수 없었다.

만약 특정한 장소에서 기다리는 택시들로 범위를 좁힌다면 택시들은 주로 공항, 시청, 터미널 등의 공공시설과 아파트 단지, 쇼핑몰, 버스 정류장과 같은 생활관련 시설들에 집중적으로 모이고 있으며 이러한 장소에서의 승객탑승 확률이 높은 편이다. 즉, 승객들이 많이 모일 수 있는 확률이 높은 곳으로서 그만큼 택시에 대한 수요가 높다는 의미가 된다. 반면 공택시들이 많이 있다면 택시의 공급이 많다는 의미가 된다. 수요가 많다면, 즉, 승객은 많은데

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

택시들의 수가 적다면 택시들이 승객을 기다리는 시간은 짧아지는 반면 승객들이 택시를 기다리는 시간은 길어진다. 반대로 택시들의 수가 많은 경우에는 택시들이 승객을 태울 때까지의 시간이 길어진다.

수요가 일정한 패턴을 유지한다고 가정한다면 공급을 수요에 맞게 유지하는 것이 택시나 승객의 대기시간을 줄이고 택시의 이윤 증대 목적에서 유리하다. 즉, 택시들이 특정 구간에 몰려서 승객을 기다릴 때 이 구간에서는 공급의 과다로 택시들의 대기시간이 커지고 그 반대의 경우엔 승객들의 대기시간이 커지게 된다. 두 경우 모두 공차율은 증가하게 되므로 택시에 대한 수요와 공급을 적정하게 유지하는 것이 필요하게 되며 이는 지역별 시간대별로 택시들을 균형있게 각 지역에 배치함을 의미한다.

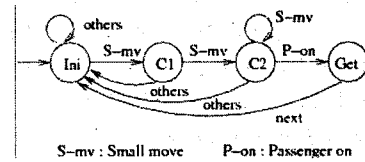
제주 택시 텔레매틱스 데이터는 지역별 수요와 공급량을 추정할 수 있는 기본 데이터를 제공한다. 즉, 전체 레코드들중 공차 상태를 갖는 보고는 택시의 공급에 해당하며 승객의 탑승에 대한 데이터는 택시의 수요량을 암시한다. 그러나 두 데이터의 스케일은 상당히 다르기 때문에 수요량과 공급량의 비교가 어렵다. 따라서 본 논문에서는 두 데이터에 대해 적절한 정규화 방법을 제시하여 각 지역에 수요와 공급을 적정하게 유지할 수 있는 배치 기법을 제시하고자 한다. 이를 위해 택시의 대기시간을 비교 기준으로 설정하고 택시의 대기시간이 가장 짧은 구간을 선택하여 이를 최적의 배치 구조라 가정하고 이 배치에 의거하여 수요와 공급이 적정하게 유지되고 있는지 판단한다.

2. 지역의 구분

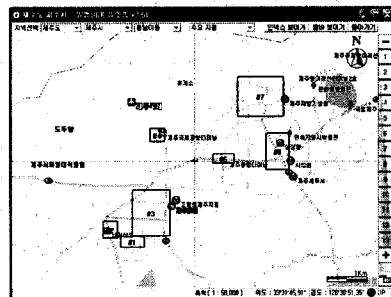
1,300,000의 레코드를 대상으로 80 % 정도가 승객을 태우지 않고 있는 상태를 보고한 반면 공차상태에서 승객이 탑승한 상태로 바뀌는 천이를 조사한 결과 8만여 회 정도이다[3]. 이 두 데이터는 각각 택시에 대한 수요와 공급을 암시할 수 있으나 그 스케일이 다르기 때문에 수요가 많

은지 공급이 많은지 비교하기 어렵다. 더욱이, 공차들의 경우 계속 이동하기도 하고 특정 지역에서 기다리기도 하는데, 이동하는 경우는 예측이나 정량화하기 불가능하기 때문에 공급량을 제시하기도 역시 불가능하다. 오직 특정 지역에서 승객을 기다리는 택시에 대해 정량화하는 것이 가능하다.

먼저 택시 텔레매틱스 시스템의 위치이력 데이터에 의해 택시들이 기다리는 지역을 추출하기 위해서는 (그림 1)과 같은 상태를 정의하여 택시들이 일정시간동안 거의 정지 상태, 즉 1분 내의 이동거리가 30m 이내인 상태(small move)로 2분 이상 지속되는 경우만을 추출하도록 한다[2]. 물론 1분 이내에 승객을 태울 수도 있지만 이는 이동중에 승객을 태운 경우와 구분할 수가 없다. 따라서 이 경우는 오차를 감수하고 대기 후 승객 탑승에서 제외하도록 한다. (그림 1)에서 각 택시 마다 하나의 상태도가 할당이 되고 각 택시의 레코드가 처음 발견되면 초기상태에서 시작하며 small move가 탐지되면 C1으로 천이한다. 결국 (그림 1)의 상태도는 2분 이상 특정지점에서 대기후에 승객을 태운 경우에 대해 위치, 시간, 대기 시간 등에 대한 정보를 생성할 수 있다.

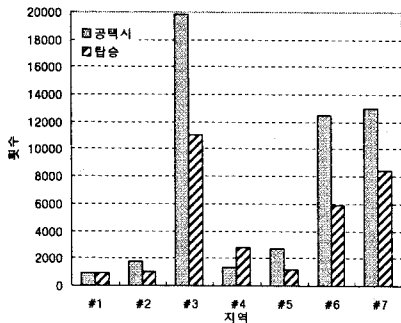


(그림 2) 대기장소 추출을 위한 상태도



(그림 3) 택시 대기 장소의 분포

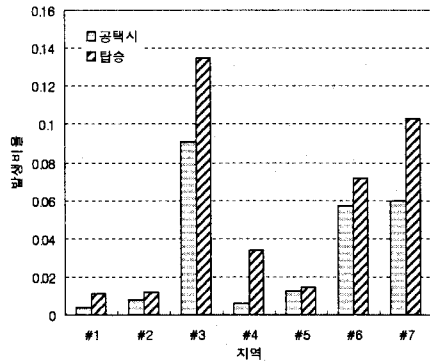
결국, (그림 1)의 상태도 분석에 의해 빈 택시들이 많이 기다리는 장소를 추출할 수 있게 되고 이중에서 다른 장소보다 현저하게 많은 탑승이 있는 장소들을 지도 상에 도시하면 (그림 2)와 같다. 이들은 반경 1.5 km 이내에 포함된다. 이 장소들은 순서대로 신제주 아파트 단지, 신제주 쇼핑 마트, 신제주 호텔 지역, 공항, 시외 버스 터미널, 시청 지역, 탑동 및 중앙로 지역 등에 해당한다. 이와 같이 분류된 위치를 기반으로 하여 각 대기장소에서 발생된 공택시 보고수와 승차 탑승 횟수를 그래프로 도시하면 (그림 3)과 같다.



(그림 4) 공택시 보고와 탑승수의 단순비교

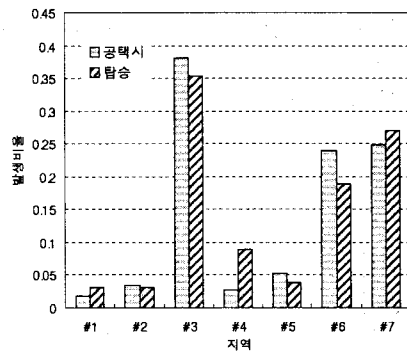
(그림 3)에서 보는 바와 같이 1,151,239 개에 이르는 공택시 보고수와 81,677 개의 탑승 수는 데이터의 양이 달라 지금 현재 어느 지역이 택시들이 많이 몰리고 있는지를 판단할 수 있는 기준이 되지 못한다. 따라서 이들을 정량화할 수 있는 방안이 필요하게 되는데 우선적으로 2 가지의 정규화 방법을 생각할 수 있다. 첫 번째 방법은 공차를 보고한 레코드 수에 대하여 지역내 보고수와 전체 보고수의 비율을 공급으로 설정하고 지역내 탑승수와 전체 탑승수의 비율을 수요로 설정하는 것으로서 이는 (그림 4)에 도시하였다. 이 방식은 수요와 공급을 각각 0에서 1.0의 구간으로 매핑하는 효과를 갖고 있기는 하지만 그림에서 보는 바와 같이 최대값이 0.14보다 작게 되며 대기장소가 아닌 곳에서의 데이터가 비율 계산에 영향을

준다. 모든 대기 지역에서 수요보다 공급이 많음을 보이고 있다.



(그림 5) 전체와 지역내 회수에 대한 비율

또다른 방법은 전체 지역이 아닌 7개 관심 지역 내에서의 보고수와 탑승수를 사용하는 것으로서 오직 대기장소들에 대한 배치에 초점을 두고 있을 때 유용하며 (그림 5)에서 보는 바와 같이 0에서 0.38까지의 구간으로 사상된다. 이 두 방법은 약간의 차이를 보이고 있으며 수요와 공급의 적정성에 대한 근거자료를 제공하지는 못하고 있다.

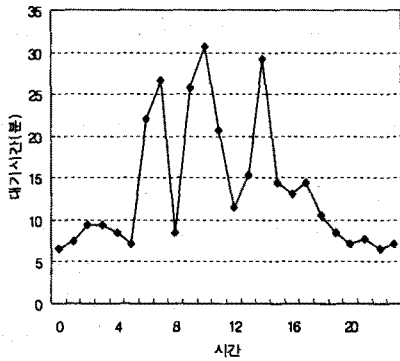


(그림 6) 지역내 회수에 대한 비율

3. 대기시간을 고려한 비교

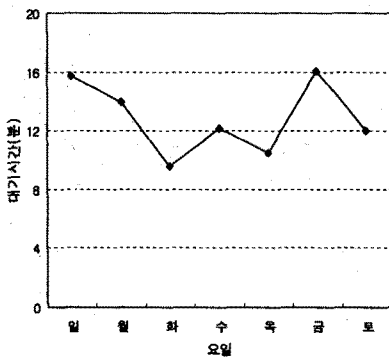
택시 텔레매틱스의 위치이력 데이터는 승객들의 대기시간에 대한 데이터는 제공할 수 없지만, 택시들이 특정장소에서 승객을 기다리는 시간에 대해 추정할 수 있는 데이터를 제공한다. 택시들의 대기시간이 짧

다면 이는 수요가 많거나 수요와 공급이 적정하게 유지되는 경우에 해당한다. 예를 들어 (그림 6)은 하루의 시간대에 따르는 택시들의 대기시간을 보이고 있는데 출퇴근 시간과 점심시간 등에 이 대기시간이 급격히 감소함을 볼 수 있다. 이 경우는 수요가 순간적으로 많아지는 경우에 해당한다. 이 그림에서 심야 새벽 시간대는 공급이 줄어 대기시간이 짧아지는 것을 반영한다.



(그림 7) 시간에 따른 대기시간

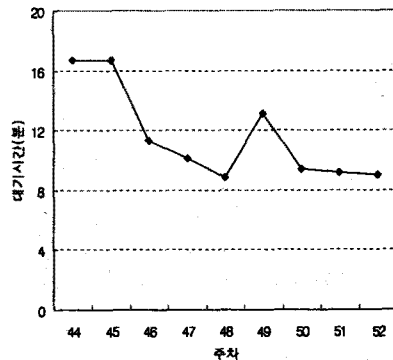
또한 (그림 7)은 요일에 따르는 대기시간을 보이고 있는데 이는 시간대에 따르는 편차를 없애고는 있으나 요일에도 예상외로 많은 영향을 받아 화요일에 수요가 많은 것을 보이고 있다.



(그림 8) 요일에 따르는 대기시간

결국 참조구간을 정할 때 위 두 방법으로

는 수요와 공급이 균형있는 구간을 찾기 어려우며 시간과 요일을 아우를 수 있는 구간으로 일주일 단위의 구간을 설정한다. 즉 (그림 8)에서 보는 바와 같이 11월부터 1월 초까지의 각 주차에 대해 8.8분의 평균대기시간을 갖는 48주차가 최소값으로서 이 시간대를 비교적 수요와 공급이 맞춰진 구간으로 간주할 수 있다. 이 구간에서 각 #1부터 #7까지의 지역에서의 공택시 비율을 48주차에서의 각 지역별 공택시 비율과 비교하여 공택시들의 위치를 조정하면 승차율을 높일 수 있을 것으로 예상된다.



(그림 9) 일주일 단위의 대기시간

4. 요약 및 결론

본 논문에서는 제주 택시 텔레매틱스 시스템의 위치 이력 데이터에서 택시사용의 수요에 해당하는 탑승 데이터와 공급에 해당하는 공택시 비율을 추출하여 수요와 공급이 적정하게 유지되도록 택시들을 배치하기 위하여 두 데이터의 스케일을 정규화하고 택시들의 승객 대기시간을 이용하여 가장 적절한 시간 구간을 발견한다. 본 논문에서 제안된 기법은 추후 보다 개선된 배치 알고리즘을 개발하는데 유용할 것으로 판단된다.

참고문헌

- [1] J. Lee, G. Park, H. Kim, Y. Yang, P. Kim, and S. Kim, "A telematics service system based on the Linux cluster," *LNCS*, Vol. 4

- 490, pp. 660-667, 2007.
- [2] J. Lee, "Analysis on the waiting time of empty taxis for the Taxi telematics system," accepted at ICCIT, 2008.
- [3] J.Lee, I. Shin, G. Park, "Analysis of the passenger pick-up pattern for taxi location recommendation," NCM, pp.199-204, Sep. 2008.