**논문 요약**

**날짜: 2020.12.02**

**이름: 전우진**

|  |  |
| --- | --- |
| 논문제목 | Origin-Destination Matrix Prediction via Graph Convolution: a New Perspective of Passenger Demand Modeling |
| 논문 요약 | 최근에는 승객들에게 편리한 승차 서비스를 제공하고 대중교통의 효율성을 향상시키는 것을 목표로 하는 디디, UCAR, 우버 등, 승차헤일링 어플리케이션들이 일상 출퇴근의 보편적인 선택이 되고 있다.  예를 들어 디디에서는 하루에 수백만 건의 택시 통화 거래가 베이징에서 발생한다[20].  고품질의 서비스를 제공하고 회사 이익을 달성하기 위해서는 승차공유 플랫폼이 승객의 요구를 실시간으로 충분히 이해할 필요가 있다.  한편, 승객의 수요를 만족시키기 위해 플랫폼은 승객들의 적시에 이동 패턴을 인지하여 서비스 차량에 대한 주문을 미리 할당해야 한다. 한편, 역사적 승객 수요로부터 대중적이고 수익성이 높은 노선을 발굴하여 빈 드라이브(즉, 승객 없이 운전하는 것)를 피함으로써 이익을 극대화하는 것이 중요하다.  따라서 단순히 지역 내에서 승객 수요의 수를 예측하는 것이 아니라, 각 여행의 출발지와 목적지에 대한 승객 수요에 대한 지식을 얻는 것이 중요하다. 서로 다른 시간대에 있는 두 지역 사이의 수요량은 승객 수요의 강도를 전달할 뿐만 아니라 유용한 이동 패턴을 채굴하기 위한 부트스트랩을 구축하기 때문이다.  다행히도, 적절한 모델링 전략과 대규모 승객 거래의 가용성으로, 이동성 패턴과 결합된 승객의 요구가 예측 가능하게 되어가고 있다.  본 논문에서는 출발지-목적지 매트릭스 예측(ODMP)으로 정의되는 새로운 관점에서 승객 수요의 모델링에 대해 조사한다.  요약하면, 출발지-목적지 매트릭스는 (1) 서로 다른 출발지와 목적지의 조합과 (2) 각 출발지 조합에 대한 승객 요구의 수라는 두 가지 정보를 포함한다.  ODMP의 목표는 주어진 시간 간격 내에 한 지리적 지역에서 다른 지리적 지역으로의 승차공유 주문 수를 예측하는 것이다. 승객들의 적시 이동성 패턴은 수요 모델링의 효과에 중요한 요소다.  예를 들어, 아침 출근 시간에는 많은 인파가 주택가에서 직장까지 이동하는 경향이 있다.  이러한 추세를 조기에 발견하고 원산지 정보로 승객의 수요를 예측할 수 있다면, 승차헤일링 플랫폼은 사전에 각 서비스 차량에 대해 가능한 배차 장소를 정확하게 제시할 수 있을 것이다.  그러나 다음과 같은 이유로 승객의 이동 패턴을 모델링하고 출발지와 목적지를 정확하게 동시에 예측하는 것은 상당히 어려운 일이다.  첫째, (1) 특정 지역의 승객 수요량과 (2) 이러한 수요의 최종 목적지를 동시에 고려해야 한다.  대부분의 수요 예측 문제는 단순한 회귀 작업으로 처리된다. 즉, 최근 특정 지역에 필요한 승차 서비스 양을 근사하게 계산한다.  그러나 이는 서비스 제공업체가 고객이 원하는 목적지를 이해하는 데 도움이 되지 않으며 인기 있는 경로에 대한 통찰력을 제공하지도 않는다.  이를 위해 승객의 이동성 기록을 그래프에 직접 모델링하는 GEML(Grid-Embedding based Multi-task Learning)이라는 모델을 제안한다.  구체적으로, 우리는 노드가 지리적 지역(그리드라고 정의됨)을 나타내고 노드 사이의 링크가 승객 수요의 존재를 나타내는 그래프로 지리적 지역과 관련된 승객의 발생 기록을 나타내며 링크의 가중치는 주문량을 나타낸다.  구체적으로, 우리는 우리의 그리드 포함 부분에 두 종류의 이웃을 고려한다. 그리고 그들은 두 개의 그리드가 지리적으로 가까운지 아니면 승객의 요구에 의해 연결되어 있는지에 기초하여 지리적인 이웃과 의미론적 이웃이다.  전자는 한 그리드와 그 인접 네트워크 사이의 본질적인 근접성을 측정하는 데 사용되는 반면, 의미 이웃은 그리드 네트워크에서 출발지와 목적지 사이의 트래픽 흐름의 의미적 강도를 모형화하는 데 사용된다.  그리드 임베딩에 의해 학습된 각 그리드의 표현을 바탕으로 승객 수요의 상당한 시간적 정보를 통합하여 ODMP를 위한 멀티 태스킹 신경망을 설계한다.  우리는 UCAR과 Didi라는 두 개의 ride-handling 애플리케이션에서 생성된 두 개의 실제 데이터셋에 대해 실험을 실시한다.  UCAR 데이터 집합은 2016년 8월 1일부터 31일까지 베이징 도심 지역에서 수집된다.  디디 데이터 집합은 2016년 11월 1일부터 30일까지이며 청두3의 도시 지역을 포함한다.  표 1은 두 데이터 집합의 특성을 요약한 것이다.  두 데이터 세트가 모두 감압되었다는 점에 유의하십시오.  우리는 표 1의 격자 세분화를 바탕으로 베이징과 청두를 각각 400과 225 그리드로 나눈다. 택시 운전사가 그러한 거리를 운전하는 데 평균 5분이 걸리기 때문에 [22]의 통계와 같이 승객들의 합리적인 대기 시간이다.  두 데이터셋의 OD 매트릭스 시퀀스는 1시간 단위로 구성된다.  광범위하게 적용되는 두 가지 메트릭, 즉 RMSE(루트 평균 제곱 오차)와 SMAPE(대칭 평균 절대 백분율 오차)를 사용하여 예측 정확도를 평가한다.  의외로 두 번째 데이터 집합에서 LSTM과 LSTNet의 결과는 분명히 HA보다 나쁘다.  LSTM의 가능한 이유는 LSTM이 ODMP 문제의 공간 속성을 무시하는 완전한 시계열 방식이기 때문일 수 있다.  반면 LSTNet은 CNN을 이용하여 공간적 특성을 모델링한다.  매트릭스에서 인접한 요소들 사이에 의미 있는 공간적 관계가 없기 때문에 OD 매트릭스에 적합하지 않다.  • 임시 그래프 데이터(GCRN 및 GEML)에 맞게 조정된 방법은 전반적인 성능을 개선한다.  이것은 우리가 ODMP 문제를 해결할 때 OD Matrix를 그래프로 모델링하는 것이 더 합리적이라는 것을 증명할 수 있다.  • 두 데이터셋 모두에서 HA의 성능은 좋지 않지만 안정적이며, 때로는 다른 기준선보다 훨씬 우수하다.  이것은 문제의 속성을 잡을 수 없다면 단순한 모델이 더 나은 선택이라는 것을 깨우쳐 줄지도 모른다. 그림 8(a)는 아침에 사람들이 집을 떠나 다른 목적지(예: 직장이나 유흥 장소)로 행진하는 경향이 있음을 나타낸다.  왜냐하면 많은 주거지역이 g202, g205, g189, g169에 위치해 있기 때문이다.  한편 g267 인근에 소프트웨어 공원이 있으며 g290은 올림픽숲 공원이다.  • 그림 8(b)에서 우리는 점심식사 후에도 여전히 많은 승객들이 일과 오락 장소로 떠나고 있음을 알 수 있다.  g168에 위치한 유명한 시단금융거리는 주거와 금융이 혼합된 지역인 g167과 g189와 함께 위치해 있다.  그리고 g253과 g233 근처에는 798 아트 지구로 잘 알려진 베이징 최대의 아트 센터가 있다.  • 그림 8(c)에 따르면 승객의 이동성에 대한 아코디언트 방향은 없다.  예를 들어, 일부 사람들은 g129에서 거의 순수한 주거 지역인 g272까지 피곤한 하루를 보낸 후에 집에 가는 것을 선호한다.  하지만 오후 9시에도 798 예술지구에 사람들이 있다.  한 가지 가능한 이유는 저녁 시간이 사람들이 원하는 목적지를 선택할 수 있는 사람들의 여가 시간이기 때문일 것이다. |
| 논문 의의 | 우리는 주어진 시간대에 주어진 기원과 목적지의 승객들의 요구를 참신하고 통일된 방법으로 예측하는 새로운 문제 ODMP를 제안한다. 이는 승차공유 플랫폼이 자동차를 준비하고 주문을 발송하는데 큰 도움이 될 수 있다.  우리는 관심 영역을 지도 상의 그리드로 나누어 ODMP 문제를 공식화한다. 그런 다음 GCNs에서 메시지 전달 스키마를 모방하여 서로 다른 그리드 간의 트래픽 전송 관계를 모델링하는 새로 정의된 그리드 이웃(지리적 및 의미적 이웃)들 간의 그래프 경합을 통해 각 그리드에 대해 임베딩을 수행하도록 그리드 내장 네트워크를 설계한다.  우리는 승객 수요의 일시적 추세를 포착하기 위해 LSTM(Long Simple-Task Memory Recurrent Networks)을 이용하는 멀티태스킹 학습 네트워크를 설계한다.  두 개의 하위 작업은 그리드에서 개별적인 들어오고 나가는 요구를 예측하는 반면, 주요 과제는 각 그리드 쌍 사이의 요구를 예측한다.  실제 및 대규모 ride-handling 데이터셋에 대한 광범위한 실험 결과, 제안된 GEML 모델이 기준선을 능가하는 것으로 확인되었다.  우리가 그리드-임베딩 프레임워크를 설계하는 이웃의 정보를 통해 지역의 이동 패턴을 모델링하는 것이다.  그리고 데이터의 범위와 넓이를 감지할 수 있도록 사전 가중치를 부여한 함수를 집계 프로세스에 추가한다. |
| 논문의 한계점 및 기타 | 특정 지역 말고 다른 지역에서도 잘 적용되는지 확인이 필요할 것 같다. |