|  |
| --- |
| **1. 주제**  딥러닝을 활용한 택시 예측 배치 시스템 개발  **분반, 팀,**  나반, 2팀, 20223082 안영찬/20223077 백승민/20223084 예준희/ 20221817 최정훈 |

|  |  |
| --- | --- |
| **2. 요약**  딥러닝을 활용하여 택시가 어디에서 몇 시에 얼마나 필요할지 예측하는 시스템을 개발하는 것이 목표이다. 수요 초과 시장에서는 항상 공급자의 의사에 따라 시장 내 자원이 비효율적으로 분배된다. 우리는 수요 초과 시장에서 자원의 비효율적 배분 현상을 완화시키기 수요값의 예측과 공급자 정보 공개가 그 열쇠라고 보았고 아마존의 물류 예측 발주 시스템에서 착안하여 수요예측 시스템을 구상했으며, 우리 팀원들이 현실에서 가장 불편하게 느끼던 택시 배차 지연 문제의 해결을 위해 적용하였다. 수요 예측을 위해 딥러닝 기술을 적용하였으며, 기술 구현을 위해 다양한 딥러닝 라이브러리르 제공하는 구글 코랩(Colab)의 환경에서 파이썬 언어를 이용하여 Sci-Kit Learn/Keas/Tensorflow 등의 인공지능 오픈소스 라이브러리를 활용하였다. 우리의 오픈소스는 학습시킬 데이터 양식만 준수한다면 “대량 자원의 수요를 예측할 필요가 있는” 대다수의 수요 초과 시장의 자원 병목 현상을 해결할 수 있고, 공급자의 정보 독점 문제를 해결책을 시사한다는 점에서 그 의의가 있다. | **3. 대표 그림**    그림 1. 대표그림 |

|  |
| --- |
| **4. 서론**  우리 조원 모두는 택시를 이용하기 위해 ‘카X오 모빌리티’와 같은 모빌리티 플랫폼을 이용한 적이 있다. 이때, 플랫폼과의 수수료나 거리 및 동선 효율 상의 이유로 택시 기사들이 승차 콜을 무시하는 경우가 다수 발생했다. 가장 효율적인 운행으로 운임을 얻어야 하는 택시기사들의 입장이 한 편으로는 이해가 가면서도, 택시를 잡기 위해 3-40분 넘게 기다리다가 결국 다른 교통수단을 이용하거나 잡더라도 약속 시간에 늦는 경우가 발생하는 게 소비자의 입장에서 너무 불편했다. 근처에 택시가 없으니 다른 교통수단을 이용하라는 정보만 확실하게 알 수 있었어도 시간-비용을 더 절약할 수 있을 것이라는 생각이 들었다. 이에 우리는 “어떻게 하면 택시들이 내 콜을 거부하지 않으면서, 빠르게 택시를 잡을 수 있을까?”라는 문제의식을 품게 되었다. 그러던 어느 날, 인터넷에서 우연히 아마존의 물류 예측 배송 시스템에 관한 자료(김보성,2021) 읽게 되었다. 이를 통해 ‘단순히 콜을 수신하고 반응하는 것으로는 배차 시간 단축에 한계가 있다고 여겼고, 물류 예측 배치 시스템처럼 택시도 거점을 두고 예측해서 배치하면 어떨까?’ 라는 생각을 품게 되었다. 이에 따라, 인공지능을 통해 합승 택시 수요 데이터를 학습시켜 해당 거점 해당 시각에 택시가 얼마나 필요할지 예측하는 프로젝트를 진행하였다.  수요 초과 시장은 늘 공급자의 정보 독점 문제와 공급 자원 병목현상으로 인해 자원이 효율적으로 배치되지 못하는 문제를 겪었다. 우리는 공간 좌표상의 거점마다 수요량을 예측할 수 있는 인공지능을 개발함으로써 수요초과 시장의 자원 병목 현상을 해결하려 한다. |

|  |
| --- |
| **5. 본론**        우리는 딥러닝 기술을 활용하여 택시 예측 배치 시스템을 개발했다. 딥러닝(천재학습백과, 2022)이란, 컴퓨터가 사람처럼 학습하고 배울 수 있도록 하는 기술로서, 인간의 축삭돌기의 학습방법에서 착안하여 인공 신경망을 구현하는 기술이다. 데이터를 입력받고 학습을 진행한 후 답을 도출하고 이를 정답과 비교해 오차를 줄여나가는 방향으로 데이터를 다시 입력해줌으로써예측 도출의 성능을 강화한다. 딥 러닝은 이렇게 인공적으로 구현한 신경 세포 ‘퍼셉트론’을 다차원 배열로 구성하여 학습한다.  텍스트이(가) 표시된 사진  자동 생성된 설명    택시 예측 배치 시스템의 개요는 위의 <시스템 개요> 사진과 같다. 먼저 소비자의 수요 데이터를 입력받는다. 실제 사용자들로부터 데이터를 입력받을 수 있었다면 좋았겠지만, 실제 데이터를 입력받기엔 시간과 자본, 기술이 부족하여 난수 데이터를 생성하고 이를 통한 학습을 진행했다. 다만, 그럼에도 AI학습의 의의를 유지하기 위해 서울시 데이터 열린 광장에서 생활인구 데이터(서울 열린데이터 광장, 2022)를 다운받고, 여기에 없는 시간 데이터, 택시 종류, 장소 데이터 등을 난수 생성기를 통해 추가했다. 추가하는 난수 데이터를 너무 무작위로 설정하면 AI가 변수 간의 패턴을 찾을 수 없기에 가중치를 부여하였고, 자연어 상태로는 AI가 학습하기 힘들어 “장소 데이터를 행정동 코드로 변환하는 등” 숫자 데이터로 변환했다.  본격적인 AI 학습을 구현하기 위해 sequential API를 구성했다. 은닉층은, ai의 학습이 일어나는 층으로 인간이 찾기 힘든 데이터 속 숨겨진 패턴을 AI가 발견하는 층이다. 그리고 원하는 뉴런의 개수, 활성화함수를 설정하여 은닉층을 추가했다. 활성화함수란 입력 신호의 총합을 출력신호를 변환하는 함수로써 은닉층이 유의미하게 작동하기 위해 필요한 함수이다. 이번 프로젝트에서 사용한 활성화 함수는 sigmoid함수와 Relu함수로 함수이다. 최적화 모듈과 손실함수는 각 은닉층들의 사이의 가중치를 조절해주는 기능으로 오차역전파를 구현하는 데 사용된다. 최적화 모듈에는 adam을, 손실함수에는 mse라는 함수를 사용했다. 이후, ai의 기본적인 설정들을 하고 100번의 반복학습을 진행했다.    정확도 80%의 예측 배치 모델을 완성했고, 학습시킨 ai에 예시로 13분, 일반택시라는 값을 넣어 보니 ‘1’ 즉, 노량진이라는 출발지를 예측해내는 것을 볼 수 있었다. |

|  |
| --- |
| **6. 결론**  Aim의 불완전한 달성이었다. 갖은 모델을 속히 말하는 “노가다”를 이용해 테스트하며 정확도를 80%까진 끌어올릴 수 있었지만, 최적화된 알고리즘을 수학적으로 모델링할 역량과 딥러닝에 관한 깊은 이해가 부족했기에 과적합이 어느 정도 발생하는 것은 막을 수 없었다. 또한, 딥러닝 모델의 태생적 한계로 인해 은닉층에서 발견된 알고리즘이 어떤 것인지 설명할 수 없다는 한계가 있다.  여러 가지 성능 및 구현 문제와 미진한 기술력에도 불구하고, 우리의 프로젝트는 예측치를 제공하는 AI의 틀을 공개한다는 점에서 그 의의가 있다. 우리의 AI가 학습할 수 있는 입력 데이터 형식만 지킨다면, 상당수의 수요 초과 시장의 문제에 대해 예측치를 제공할 수 있게 된다. 예를 들어 야구에서는 투수의 투구 패턴을 분석해 다음에 올 공의 종류와 위치를 예측하여 타격 성적을 상승시킬 수 있을 것이고, 옥외 광고 분야에서는 유동 인구와 시간대를 예측하여 원하는 시청자 층에게 타깃으로 광고할 수 있을 것이다. 우리는 우리의 프로젝트가 예측 데이터를 필요로 하는 모든 분야에 간단한 실험실을 제공할 수 있다는 점에서 그 의의가 있다고 본다.  향후의 연구는 난수 데이터 대신 카카오 모빌리티 같은 다수의 사용자를 가진 플랫폼과 협력하여 실제 사람들로부터 데이터를 받아 유의미한 가치를 창출하는 방향으로 진행해야 한다. 또한, 딥러닝을 더 깊게 공부하여 최적화된 딥러닝 모델을 수학적으로 설계하고, 과적합현상을 방지해야 한다. 그리고, 시스템 그 자체로는 택시 수요자들과 기사들이 편하게 정보를 인식하고 제공할 수 없기 때문에 사용자들이 이용하기 편리하게 앱이나 웹페이지를 만들어야 한다. 마지막으로, 실제 수요자들은 다양한 곳에서 택시를 수요하므로, 수요 데이터를 실제 gps상의 데이터로 받고 이를 군집화하여 분류하는 알고리즘도 설계하여 택시 수요자들을 거점으로 오게 하지 않고, 자리에서 즉시 택시 서비스를 이용할 수 있도록 진행되어야 한다.  -소스 코드 배포  Github로 공유함.  <https://github.com/Latuca/Open_Source_Young.A> |

**7. 출처**

[1] 김보성, 2021, 로티스 홈페이지, “아마존의 예측배송”, 물류기술동향, <https://www.lotis.or.kr/trends/3030>

[2] 천재학습백과, 2022, <https://terms.naver.com/entry.naver?docId=3609903&cid=58598&categoryId=59316>

[3] 서울 열린 데이터 광장, 2022, <https://data.seoul.go.kr/dataList/OA-14991/S/1/datasetView.do>

[4] 심지환, 2022, [[Lecture\_3] 머신러닝의 이해 및 실습.pdf](file:///C:\Users\AA\OneDrive\ë°í%20íë©´\AI_Engineer_ìì±ê³¼ì \3ì¼ì°¨%20-%20ë¨¸ì ë¬ëì%20ì´í´%20ë°%20ì¤ìµ\%5bLecture_3%5d%20ë¨¸ì ë¬ëì%20ì´í´%20ë°%20ì¤ìµ.pdf) , 국민대학교 AI빅데이터 경영학과

[5] 심지환, 2022, [[Lecture\_4] 딥러닝의 이해 및 실습.pdf](file:///C:\Users\AA\OneDrive\ë°í%20íë©´\AI_Engineer_ìì±ê³¼ì \4ì¼ì°¨%20-%20ë¥ë¬ëì%20ì´í´%20ë°%20ì¤ìµ\%5bLecture_4%5d%20ë¥ë¬ëì%20ì´í´%20ë°%20ì¤ìµ.pdf) , 국민대학교 AI빅데이터 경영학과