

Page 1)

안녕하세요? 저희는 이번 2022년 인공지능과 공간정보처리 집중 교육 및 경진대회에 참여한 컴퓨터교육과 양래은, 전산통계학과 박지현, 윤소원입니다.

Page 2) 팀 소개

저희 팀은 전산통계학과 Cyber Physical System에서 알고리즘과 다양한 데이터 분석에 대해 공부하고 있는 그룹입니다. 저희는 제주지역에서 발생하는 데이터에 대해 관심이 많으며 알고리즘과 인공지능을 활용한 데이터 분석 분야로 취업을 하고자 노력중입니다. 작년부터 제주지역에서 발생하는 전기자동차 충전기 상태 데이터를 수집하고 있는데 단순한 파일 다운로드가 아닌 OpenAPI를 통한 주기적 크롤링 방식을 사용하여 데이터 컬렉터를 구현하였습니다. 이 데이터의 활용방안을 고려하고 있던 중에 이번 교육 프로그램에 대해 알게 되었고 인공지능에 관련된 분석 기법 등을 배우고자 프로그램에 지원하였습니다.

Page 3) 개요

저희 팀이 이번 과정을 통해 진행한 바는 데이터의 자동 수집과 데이터베이스 저장, SQL 질의를 통한 데이터 가공, 파이썬 개발환경의 구축, 파이썬의 다양한 데이터분석, 그래프 라이브러리를 활용한 예측 모델의 개발 등입니다. 사실 취업능력 배양이 궁극적인 목표이긴 하지만 제주지역 전기자동차 사용률에 대한 분석을 지속적으로 발전시켜 신재생 에너지와의 결합, 정책결정을 위한 데이터모델 등을 개발하고자 합니다.

Page 4) 데이터 수집 구조

먼저 저희가 구현한 데이터 크롤러를 수행시키려면 공개 데이터 사이트에 회원으로 가입하고 한국환경공단에서 제공하는 전기자

동차 충전소 정보 조회 데이터셋에 대한 사용허가를 받아야 합니다. 허가는 문자열 키로 주어지는데 파이썬 코드에서 이 키를 내장 시켜 Restful API를 연결합니다. 다음 CPS 연구실에 있는 리눅스 서버에서 크롤링 코드를 수행시키는데 주기적인 수행을 위해서 크론탭에 프로그램을 등록하여야 하며, 이 데이터를 MySQL 데이터베이스에 일차적으로 자동 저장합니다. 따라서 기존의 서비스와 같이 단순히 어떤 시점의 상태만 보여주는 게 아니라, 타임스탬프가 부여된 연속 데이터 스트림을 생성하고 있습니다.

Page 5) 데이터 수집 코드

이는 저희가 작성한 파이썬 코드의 일부입니다. 사용허가 후 부여된 문자열 키를 내장한 Retrieve, BeautifulSoup parser와 MySQL 명령을 생성하는 부분도 보입니다.

어디..?

Page 6) 데이터 베이스 입력

결국 자동적으로 Linux 서버의 데이터베이스에 충전기의 위치, 타입 등 정적정보를 가진 charger 테이블과 주기적으로 계속 쌓이는 OpStatus 두 테이블이 생성되었습니다. OpStatus에는 지금도 데이터가 쌓이고 있습니다. 이 화면은 저희 리눅스 서버의 터미널에서 캡처한 것입니다. + charger와 OpStatus 차이..?

Page 7) 사용률 분석과 지도 플로팅

저희는 이번 강좌에 포함된 Shape 파일에 기반한 데이터 분석까지는 좀 미치지 못했지만 파이썬의 Folium 라이브러리를 이용하여 충전기들의 위치를 지도상에 플로팅하였습니다. 뒤에 설명이 되겠지만 각 충전기들의 사용율을 계산하여 레벨별로 다른 색을 표시하였습니다. 모든 좌표는 WGS84 좌표계를 따릅니다. 대체적으로 사람이 적은 외곽보다 사람이 많은 시내에 충전기가 몰려 있는 것을 알 수 있습니다. 사용률을 바탕으로 어느 지역에 충전기를 추가 설치할지 어느 지역 충전기에 고장이 의심스러운지 등을 파악할 수 있습니다.

→ 방문?
WGS 좌표

Page 8) 관련 코드

지도를 플로팅하기 위해 작성한 코드입니다. SQL 코드로 각 충전기별 유효한 상태정보 레코드 수와 그 중에서 충전중이라고 표시된 레코드 수의 비율을 사용률로 계산합니다. Foilium 라이브러리로 지도를 불러온 후 사용률 레벨에 맞는 색으로 아이콘을 표시했습니다.

Page 9) 충전기 전력 정보

참고적으로 각 충전기 타입에 따른 사용률을 파악할 수 있는데 완속 충전기에 비해 급속 충전기들이 높은 사용률을 보이고 있습니다.

Page 10) SQL 코드 부분

사용률을 계산하는 SQL 코드입니다. 질의문에서 레코드들을 조회하는 조건을 다양하게 변화시키면서 수행시켜 결과를 텍스트 파일에 저장하고 파이썬 프로그램에서 처리하도록 하였습니다.

Page 11) 4점 데이터의 신뢰성

Data.go.kr에서 공개하고 있는데 데이터의 신뢰성을 파악하기 위해 각 날짜별로 수집된 유효한 상태정보의 수를 플로팅한 결과입니다. 저희 리눅스에서 프로그램은 고장없이 잘 수행되었지만 아마도 환경공단 서버에서 데이터를 공시할 때 5일 정도 문제가 있었던 것으로 보입니다. 일별 사용률은 충전이 거의 없는 심야시간까지 포함해서 20% 정도를 보이고 있습니다.

Page 12) 주변 충전기와의 사용률 연관성

직관적으로 충전기별 사용률을 보면 충전기가 밀집된 지역에 수요가 몰린 것을 알 수 있었습니다. 가까운 곳에 다른 충전기가 있으면 수요가 분산될지 아니면 해당 충전기군으로 더욱 수요가 몰릴지를 파악하고자 하였습니다. 제주시청과 같은 공공 충전장소에는 충전기가 10기 이상 있는 곳도 있고 대부분 2개 이상의 충전기들이 설치되어 있습니다. 가장 가까운 거리의 충전기와 사용률 간의 상관관계는 0.54 정도로 일단 무시할 수는 없는 수준입니다. 충전기간의 거리를 구하는데 있어서는 도로망 정보를 구할 수 있

구 어떻게 구했는지.

다면 다익스트라 알고리즘에 의해 네트워크 거리를 구할 수 있지만 그 정보가 없어서 **유클리드 거리**를 계산하였습니다.



Page 13) **주기성 분석**

이 그래프는 주기성 분석 결과입니다. 크진 않지만 일주일 단위의 주기성이 있는 것으로 보여 주기분석을 해본 결과로서 ~~trend~~, ~~주~~ **관측치**, ~~기성~~, 잔차 등의 결과가 보여집니다. 저희가 추적하고 있는 데이터는 아직 2달치 정도라 전반적인 trend를 파악하기엔 부족합니다. 좀더 데이터가 쌓이면 제주도 전반적인 수요 패턴에 있어서의 추세를 발견할 수 있을 것이라 생각합니다.

추세, 계절성, 잔차

Page 14) **관련 코드**

주기성 분석을 수행하고 플로팅하는 코드의 일부입니다.

Page 15) **시간과 요일별 사용률**

시간대별 사용률을 플로팅한 결과 10시부터 17시까지 사용률이 높습니다. 전력수요가 많은 시간대에 충전이 많이 일어나고 있어서 **이 시간대에 신재생 에너지와의 연계가 필요합니다**. 오른쪽 그림은 요일별 사용률인데 앞에서 언급한 바와 같이 요일별 사용률에는 큰 차이가 없으며 일요일과 월요일에 약간 낮은 걸 알 수 있습니다.

Page 16) **충전기별 사용률**

충전기별 사용률인데 충전기들 간에 많은 편차가 있는 것을 알 수 있습니다. 대부분의 충전기들이 20% 정도의 시간만 충전에 사용되는데 일부 충전기는 60% 정도까지 사용되고 있으며 이 지역에 증설이 필요할 것으로 보입니다. **분산가능한 수요인지 파악하기 위해서는 충전자들의 특성을 알 수 있어야 합니다.** ??..

Page 17) 「

이 그래프는 충전기 마다 1시간 단위로 시간슬롯을 나눴을 때 피크 시간대에는 어느 정도의 **부하**가 물리는지 측정한 결과입니다. 역시 극심한 사용률 편차를 보여주고 있으며 일부 충전기들은 특

→ 기계적 충위 소비에너지 크기

정시간대에는 80 % 이상 점유되어 있어서 충전을 원하는 사용자들은 해당시간에는 그 충전기의 사용을 시도할 필요가 없습니다. 이런 데이터를 충전기 사용자에게 제공한다면 부하를 일부 분산할 수 있을 것으로 보입니다.

Page 18) 인공신경망에 의한 예측

다음 그래프는 사용률, 혹은 충전 수요에 대한 예측 모델을 파이썬의 인공신경망 라이브러리를 이용하여 구축하는 과정을 보이고 있습니다. 출퇴근처럼 주행거리가 일정한 충전하는 사람들 입장에서 보면 아주 타이트하진 않지만 며칠마다 한번씩 충전을 하기 때문에 전체적으로도 주기성이 있을 것으로 보입니다. 또, 행사나 계절적 요인으로 시간구간의 사용률이 높아지거나 낮아질 수 있을 것으로 예상됩니다. 따라서 우선적으로 이를 시계열 데이터로 간주하여 지난 4일 동안의 사용률에 따라 내일의 사용률을 예측하여 보았습니다. 여기서 요일 요소는 모델에 넣기도 하고 빼기도 하였습니다.

→ 제품 판매량 예측

Page 19) //

먼저 저희가 수집한 데이터를 쿡킹하여 기계학습 패턴을 생성하였습니다. 파이썬의 사이킷런 라이브러리를 이용하여 데이터의 70%는 학습에, 30%는 예측에 적용하였고 그 결과 피팅과 예측이 섞인 그래프가 보이고 있습니다. 왼쪽 그림은 요일 요소를 넣지 않은 모델, 오른쪽은 요일 요소가 포함된 모델의 결과입니다. Root mean square 값이 사용률의 10 % 수준인 0.026 정도에 머물고 있으며 역시 데이터가 좀더 쌓이면 모델의 타당성을 알 수 있을 것으로 예상합니다. 또, 지금은 기후 요소를 포함시키지 못했는데 기상청 사이트에서 데이터를 다운로드받아 모델에 결합시킬 계획입니다.

다층신경망

Multi-Layer Perceptron

Page 20) 관련 코드 (일부)

이에 관련된 코드의 일부입니다. MLP 등에 관련된 함수들이 보입니다.

결론적으로 저희는 이번 프로그램을 통해 파이썬을 이용한 인공지능 및 빅데이터 처리 기술에 대한 이해를 넓힐 수 있었고 특히 파이썬의 기계학습 라이브러리들을 활용할 수 있게 되었습니다. 제주지역에서 발생하는 데이터에 관심을 갖게 되었는데 향후 교통, 환경, 에너지 등 다양한 분야에서 생성되는 데이터들은 무한한 가치를 갖고 있다고 생각합니다. 저희의 목표는 당연히 취업인데 이를 위해 데이터들을 활용하는 경험을 꾸준히 가질 것이며 알고리즘 등 여러 경진대회에 참여할 예정입니다. 이상으로 발표를 마치겠습니다. 감사합니다.