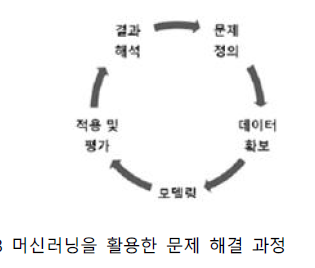
3월1일~3월6일

3월1주차 졸업논문 주제 선정 및 개념공부

가장 관심있는 분야인 머신러닝을 큰 틀을 잡고 주제를 선정할 것입니다.

-머신러닝

machine learning이란 일반적으로 생성된 데이터를 지적으로 사용할 수 있는 정보로 변환하는 컴퓨터 알고리즘을 말한다. machine learning은 1950년대 후반 AI(인공 지능) 기술로 도입됐다. 시간이 지남에 따라, 그 초점은 진화하고 계산적으로 실행 가능하고 강력한 알고리즘으로 전환되었다. 지난 10년 동안, machine learning 기술은 생물 정보학, 음성 인식, 스팸 탐지 및 광고 네트워크와 같은 다양한 응용 분야에서 분류, 회귀 및 밀도 추정을 포함한 광범위한 작업에 사용되었다. 알고리즘 및 기술은 통계, 수학, 신경 과학 및 컴퓨터 과학을 포함하여 IoT(internet of Things)의 다양한 분야에 대한 최근의 기계 학습 응용 프로그램에서 왔다. 이러한 machine learning 개념은 빅데이터 혹은 데이터 마이닝과 혼동이 되는 경우가 종종 있다. 데이터 마이닝은 많은 데이터 속에서 유효한 정보를 찾아내는 과정에 가깝고, 머신러닝은 데이터를 이용하여 검증과 학습의 과정을 통해 특정 조건에서 예측 값을 얻는 과정이라는 점에서 차이가 있다. 기계학습의 가장 기본적인 절차는 데이터를 정제하고 정리하는 과정을 통해 추상화한다. 추상화된 모형을 기반으로 훈련과 데스트를 통해 일반화된 알고리즘을 도출하는 과정이다. 훈련과 테스트라는 과정을 빼고 생각해 보면 일반적인 기업 전략분석(BI Business Intelligent)시스템의 ETS(Extract Transform Load)과 비슷해 보인다. 하지만 ETL은 훈련과 알고리즘의 선택에서 차이가 있기 때문에 데이터의 정제와 변화과정에 있어서도 머신러닝과는 다른 방식으로 사용하게 된다.

machine learning을 통한 문제 해결 과정은 일반적으로 Figure 3과 같이 5단계로 정의될 수 있다. 첫 번째 단계에서는 machine learning의 적용을 위해 목표에 대한 문제를 명확하게 정의하고, 적절한 machine learning 방법론을 선택한다. 정의된 문제와 분석 방법론에 따라서 필요한 데이터를 수집하고 통합하며, 탐색 및 전처리 작업을 거쳐 분석 가능한 형태로 변환한다. 다음으로 machine learning 방법론을 활용하여 데이터에 대해서 모델링을 수행하고, 그 결과로 얻어진 모델을 실제 문제 해결에 적용하고 성능을 평가한다. 마지막으로 적용 결과를 해석하여 의미 있는 지식을 도출하며, 그에 따라서 추가적으로 새로운 문제를 정의하고 앞의 문제 해결과정을 반복한다. 여러 번의 반복적인 과정을 통해서 문제 해결 목표에 대한 개선된 결과를 도출할 수 있다.