



머신러닝/딥러닝 학습로드맵 세미나

발표자 : 홍석준(의료 스타트업 AI 개발자)

주요 경력

- 연세대 법대, 경제학(부전공)
- 김앤장 법률사무소 기획 경력
- 이랜드 전략기획본부 기획 경력
- 영국 리즈 데이터 사이언스 석사

1. 머신러닝 딥러닝 트렌드

Tools 트렌드

파이썬 : 머신러닝에서 가장 활발하게 사용되는 프로그래밍 언어(압도적)

딥러닝의 현재 트렌드는 PyTorch가 가장 높은 성장세를 보이는 중

딥러닝 분야 트렌드

자연어 처리 - 딥러닝 모델인 BERT를 이용한 자연어 처리 분야가 떠오르는 중

BERT를 통한 자연어 처리가 떠오르는 이유

- **전이학습**을 통한 딥러닝 모델이 좋은 성과를 많이 보이고 있는 현재 BERT가 이런 트렌드를 이끌고 있기 때문
- BERT가 Self-supervised Learning 모델인데 실무에서는 답이 존재하는 데이터의 양이 매우 적기 때문에 이러한 상황에 가장 정확함 방법론이기 때문에 트렌드로 자리매김

머신러닝의 다른 트렌드

eXplainable AI (블랙박스 모델을 해석하려는 분야)

대표적인 예 : SHAP

의료분야에서 특히 많이 사용

2.머신러닝/딥러닝 학습 방향

인공지능 - 규칙 기반 시스템 + 머신러닝

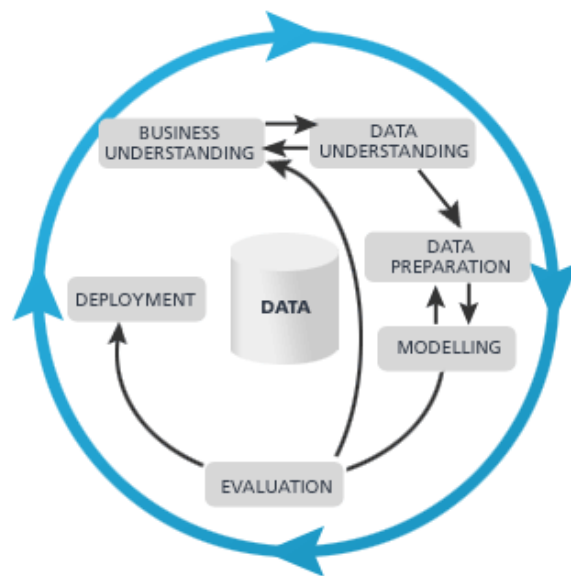
머신러닝 - 데이터에서 기계가 스스로 패턴을 학습함

딥러닝 - 인공신경망 알고리즘을 사용(자연어처리, 이미지처리 큰 성과)

데이터 사이언스의 역량

- 도메인 지식
- 컴퓨터 공학
- 통계

*CRISP-Data Mining의 프로세스를 따라야 하기 때문에 위의 세 가지 역량을 갖춰야 함.



머신러닝 공부 방향 : 지도학습 → 전이학습 → Self-supervised Learning

*이후의 학습 방향은 개인적으로 비지도학습 → 준지도학습 → 강화학습을 추천하심.

3.머신러닝/딥러닝 학습 로드맵(이론, 기술 상세 설명)

1.프로그래밍 언어 (Python)

↓

2.데이터 이해 & 전처리 (Numpy, Pandas...)

↓

3.머신러닝 기초 (Sklearn, Tensorflow(Keras)/Pytorch...)

↓

4.머신러닝 심화 (Ensemble, Self-Supervised Learning, 전이학습...)

공통적인 부분

- 수학(선형대수, 미적분, 확률론)
- 데이터 수집
- 데이터베이스
- 빅데이터 시스템(Hadoop/Spark)

※취업의 경우에는 3번 단계까지만 해도 가능한 하지만 선택 폭이 매우 좁음

머신러닝 기초 세부화

- 전통적인 머신러닝(선형회귀, KNN, Decision Tree, SVM...)

↓ (심화)

Ensemble(Bagging, Boosting)

- 딥러닝 지도학습(CNN, RNN, LSTM...)

↓ (심화)

Self-Supervised Learning(BERT, GPT...) & 전이학습(ResNet, VGG Net...)

4. Q&A

Q1.

인공지능 직무에 대한 대략적인 분류 및 설명(정보보안 업계에서 보안관제, 모의해킹, 취약점 진단, 컨설팅 등으로 업무 분류되어 있는거처럼 인공지능분야에서는 어떤식으로 직무가 분류되어 있는지)

A1.

- 데이터 엔지니어 - 빅데이터를 유실되지 않고 저장할 수 있도록 가공하는 역할(Hadoop)
- AI 개발자

- ML OPS
- 데이터 사이언티스트
- 데이터 분석가
-
-

Q2.

입문자가 공부하기에 개발환경은 어떤os가 가장 적합할까요?

A2.

Window를 추천한다. 물론 실무에 가면 Linux를 많이 이용하기 때문에 어느정도 적응을 한 후에는 Linux를 익히기 위해 넘어가는 것을 추천

Q3.

AI 결과에 대해 설명하기가 항상 힘들었는데 eXplainable AI 에 대해 추가 설명 부탁드립니다. 전부 수학이겠죠?

A3.

eXplainable AI는 대표적으로 SHAP가 있는데 블랙박스 모델을 Local하게라도 해석이 가능하게 만들어 주는 기법이며 라이브러리가 존재하기 때문에 사용하는 것은 어렵지 않으며 깊게 들어가면 수학적인 것은 맞지만 이런 부분까지 세세하게 알아야 할 필요는 전혀 없습니다

Q4.

데이터사이언스 석사학위에 대해서 관심이 많은데 한국의 석사학위와 영국에서의 석사학위를 현업에서는 똑같이 바라보나요?

A4.

큰 차이는 없으며 어떤 과목을 배웠고 기업의 실무와 연관성이 높은지가 더 중요하다.

Q5.

비전공자로 머신러닝을 배워 취업을 했는데 수학적인 지식이 부족해 어떻게 돌아가는지 잘 몰라서 항상 불안한 마음이 큰데 어떻게 공부를 해야할까요??

A5.

선형대수가 대부분의 비중을 차지하기 때문에 먼저 공부하는 것을 추천하며 개인적으로 추천하는 강의는 **한양대학교 이상화 교수님의 선형대수학 강의를** 추천한다.

Q6.

pandas, matplotlib 같은 걸로 데이터를 분석하는 걸 빅데이터 혹은 데이터분석 업무를 한다고 보고 기계학습으로 분석하는 행위 하는 걸 인공지능이라고 단순 분류하면 될까요? 그리고 취업하려고 보니 대체로 저 둘을 확실히 분류해서 채용한다기 보다는 그냥 하나로 통쳐서 데이터분석 또는 인공지능 직무로 채용하는 거 같던데 맞나요?

A6.

이부분은 특히나 한국의 경우가 통합해서 설명하는 경우가 많다. 현재는 앞서 설명한 pandas 같은 것만 사용해서 실무에서 분석을 하는 경우가 사라지고 있기 때문에 이런 현상이 일어나고 있는 것 같다. 실제로 머신러닝이 분석과정에 빠질 수 없기 때문에 통합의 개념으로 생각해서도 좋을 것 같다.
