E-Business 개론  
Prof. 조준서  
201003261 조민철

과제#1 기계학습(machine learning)과 딥러닝(Deep learning)

(10pages)

|  |  |
| --- | --- |
| **Q1. 기계학습 / 딥러닝이란 무엇인가?** |  |

**<기계학습과 딥러닝의 정의>**

기계학습(machine learning)이란 컴퓨터가 주어진 데이터를 반복적으로 분석해 의미를 찾아내고 미래를 예측하는 능력을 뜻한다. 딥러닝(Deep learning)은 기계학습 분야의 하나로서, 수많은 데이터를 이용하여 사람만이 할 수 있던 일들을 컴퓨터에게 학습시키는 것이다. 즉, 사람이 지식과 경험을 바탕으로 문제 상황을 해결하는 능력, 방대한 자료를 분석해 스스로 의미를 찾는 학습능력, 그리고 시각 및 음석인식 등의 지각능력, 자연언어(natural language)를 이해하는 능력과 자율적으로 움직이는 능력 등을 컴퓨터로 실현하는 기술이다.

**<기계학습의 현재 위상과 예시>**

먼저 딥러닝의 상위 카테고리인 기계학습부터 살펴보자. 기계학습은 인공지능 분야에서 개발된 것이며 이제껏 구현되지 않았던 컴퓨터의 새로운 능력이라고 간주된다.

**예시#1 데이터베이스 마이닝**

기계학습의 예시로는 데이터베이스 마이닝(Database mining)이 있다. 기계학습이 발전하게 된 이유 중 하나는 웹과 모든 것이 자동화되었기 때문인데, 이는 전보다 우리가 가지고 있는 데이터가 훨씬 많아졌기 때문이다. 현재 주목 받고 있는 주요 마이닝 데이터는 아래와 같다.

1) 클릭스트림 데이터

예를 들어, 실리콘밸리의 수많은 회사들이 웹에서 클릭된 데이터를 수집하려고 한다. 이를 클릭스트림 데이터(Clickstream data)라고 하는데 기계학습 알고리즘을 통해 이 데이터들을 수집하면 사용자를 더 잘 이해하고 그들에게 더 나은 서비스를 제공할 수 있기 때문이다.

2) 의료 기록

환자의 의료 기록 또한 자동화되면서 전자 의료기록부 데이터가 생겨났고 의료 기록들의 분석을 통해 환자의 병을 낫게 하는 데 도움이 되고 있다.

3) 생물학

생물학 계산 자동화를 통해 생물학자들은 엄청난 양의 유전자를 분석할 수 있게 되었고 기계학습 알고리즘이 인간 게놈을 이해하는 데 도움을 주게 되었다.

**예시#2 자동화 학습 프로그래밍**

사람이 손으로 프로그래밍을 할 수 없을 때 사용하는 것이 자동화 학습 프로그래밍(Applications can’t program by hand)이다.

1) 자동 비행 헬리콥터

자동 비행 헬리콥터를 사람이 프로그래밍 코드를 짜서 만드는 것은 매우 어렵다. 하지만 컴퓨터가 스스로 학습하도록 하면 가능하다.

2) 필체 인식

미국에서 해외로 우편물을 배송할 때 비용이 많이 들지 않게 하는 방법으로 현재 쓰이고 있다. 편지봉투에 써 놓은 글씨를 학습 알고리즘이 손 글씨를 분석하여 자동으로 분류한다.

3) 자연어처리(NLP: Natural Language Process)

컴퓨터가 언어를 이해하는 것이다.

4) 컴퓨터 비전(Computer Vision)

컴퓨터가 사진이나 그림과 같은 이미지를 인식하는 것이다.

**예시#3 개인화 프로그램(Self-customizing programs)**

스스로 프로그래밍할 수 있는 학습 알고리즘이 널리 사용되고 있다. 1) 아마존 책 추천 시스템, 2) 넷플릭스 영화 추천 시스템은 수많은 사용자들에게 맞는 추천을 하기 위해 개별적으로 프로그래밍을 할 수는 없다. 기계가 스스로 추천을 할 수 있게끔 만드는 것이 유일한 방법이다.

**<딥러닝의 원리>**

1) 분류

딥러닝의 기초적인 작동방식은 분류(classification)에 있다. 분류라는 것은, 데이터를 입력하고, 입력한 데이터에 그에 해당하는 레이블(labe: 식별기호)을 부여하는 것을 뜻한다. 가령 손으로 쓴 편지에서 알파벳 기호를 인식할 때 A처럼 보이는 것을 A로, B로 보이는 것을 B로 레이블을 부여하는 것이 바로 분류이다.

2) 트레이닝 데이터 세트(training data set)

트레이닝 데이터 세트란 이미 분류된 데이터이다. A처럼 보이는 것을 A로, B처럼 보이는 것을 B로 이미 레이블이 분류된 데이터 세트를 의미한다.

3) 테스트 데이터 세트(test data set)

우리가 풀어야 할 바로 그 문제에 해당한다. 즉 분류해야 할 데이터를 뜻한다.

**<분류 알고리즘 예시>**

분류는 딥러닝 뿐만 아니라 기계학습을 가능하게 하는 기초 원리이다. 분류를 할 수 있다면 감지(detect)하거나 순위를 매길(rank) 수 있다. 감지를 이용한 예는 다음과 같다. 당신의 차에는 도로 위 상황을 찍고 있는 카메라가 있다. 당신은 차 앞에 보행자가 있는 지 감지하여 자동차 사고를 방지하고자 한다. 이때 어떤 방식으로 보행자를 분류할 수 있는가?

보행자가 맞는 경우와 보행자가 아닌 경우인 이항분류(binary)로 분류 케이스(classifier)를 나눠야 한다. 가령 보행자가 맞는 경우에는 “pedestrian”, 보행자가 아닌 경우 “no pedestrian”이라는 레이블을 부여하고 수많은 데이터를 컴퓨터에 입력시켜서 학습시키면 된다.

순위를 매기는 랭킹 시스템의 예시로는 웹 검색엔진이 있다. 당신은 특정 쿼리(query: 검색어)를 입력하여 웹 상의 모든 문서에서 그 쿼리와 관련이 있는 페이지를 찾고자 한다. 이때 분류 케이스(classifier)를 어떻게 사용해야 하는가?

쿼리와 웹 페이지를 하나의 쌍으로 두고, 관련이 있는 경우에는 “relevant”, 관련이 없는 경우에는 “not relevant”로 레이블을 부여하면 된다.

<로지스틱 분류(logistic classifier, or linear classifier)>

-