7장 규칙기반 인공지능

개념

1. 논리: 다양한 연산을 통해 규칙기반 인공지능에서 폭넓게 활용되고 있다. 인공지능에서 지식의 표현이나 추론의 도구로 많이 쓰고 있다.
2. 명제 논리: 주어와 술어를 구분하지 않고 전체를 하나의 식으로 처리하여 참 또는 거짓을 판별하는 법칙
3. 술어 논리: 주어와 술어로 구분하여 참 또는 거짓에 관한 법칙
4. 명제: 어떤 사고를 나타내는 문장 중 애매하지 않고 명확하게 참이나 거짓을 구분할 수 있는 문장이나 수식
5. 명제의 진리값: 명제가 참 또는 거짓의 값을 지닐 때 그 값
6. 추론: 이미 알려진 사실이나 명제를 근거로 삼아 미지의 사실에 대한 판단이나 결론을 이끌어내는 사고 과정
7. 알고리즘: 주어진 문제의 해결을 위해 필요한 여러 가지 단계들을 체계적으로 명시해 놓은 것
8. 휴리스틱: 어떤 사안 또는 상황에 대해 엄밀한 분석에 의하기보다 비슷한 문제에 대한 과거의 경험들을 바탕으로 직관적으로 판단하여 선택하는 의사결정 방식

용어

1. Prolog: 추론에서 탁월한 속도와 성능을 나타낸다
2. 삼단논법: 미리 알려진 두 판단으로부터 그들과는 다른 하나의 새로운 판단으로 이끄는 추론 방법
3. 순방향 추론: 출발 상태로부터 목표 상태를 향해 진행하는 기법
4. 역방향 추론: 주어진 목표로부터 처음 출발 단계까지 거꾸로 진행
5. 지식베이스: 알려진 지식을 모아 놓는 것
6. 깊이 우선 탐색(DFS): 첫 정점을 방문하고 왼쪽으로 이동하여 계속 탐색한다. 탐색할 정점이 없으면 되돌아와서 다른 정점들을 순환적으로 탐색한다
7. 너비 우선 탐색(BFS): 시작 정점을 방문한 후 시작 정점과 연결된 모든 정점을 왼쪽부터 차례대로 방문한다. 그 후 level 순서에 따라 차례로 탐색한다
8. 휴리스틱 탐색: 앞의 길이 우선 탐색과 너비 우선 탐색은 어떠한 사전 정보를 사용하지 않고 일정한 순서대로만 탐색하는 무정보 탐색이다. 이와는 달리 탐색 과정에서 경험적 지식을 활용하는 것이 휴리스틱 탐색이다
9. 최소최대 탐색: 나의 가능성을 최대로 하고 상대방의 가능성을 최소로 하는 전략 탐색
10. 최상우선 탐색: 휴리스틱 탐색에서 목표에 도달할 가능성이 가장 큰 방향으로 탐색하는 것
11. A\* 알고리즘: 출발점으로부터 목표지점까지의 최적 경로 탐색의 한 방법으로서 적절한 휴리스틱을 사용하여 최적화할 수 있는 알고리즘
12. 지식처리 기술: 규칙이나 프레임에 의한 표현이 실용화되어 있다
13. 규칙기반 시스템: 사실을 나타내는 지식베이스와 추론 규칙으로 이루어진다
14. 전문가 시스템: 어떤 특정 분야에서의 인간 전문가들의 전문 지식을 수집 정리하여 주어진 특정 전문영역에 관한 문제를 인공지능의 추론 능력을 이용하여 해결하고자 하는 컴퓨터 자문 시스템의 일종

이해

1. 인공지능에서 많이 쓰이는 논리 연산
   1. 논리 부정: 임의의 명제 p가 주어졌을 때 그 명제에 대한 부정
   2. 논리합: 임의의 두 명제 p, q가 ‘또는(OR)’으로 연결되어 있을 때
   3. 논리곱: 임의의 두 명제 p, q가 ‘그리고(AND)’로 연결되어 있을 때
   4. 조건: 임의의 명제 p, q의 조건 연산자는 ‘p이면 q이다’라고 읽는다
2. 추론의 종류
   1. 귀납 추론: 개별적인 사실들을 분석하고 이로부터 큰 범주에 대한 일반적인 결론을 끌어내는 방법
   2. 연역 추론: 기본적으로 하나 이상의 전제로부터 다른 어떠한 결론을 도출해내는 추론 방식
   3. 유비 추론: 두 개의 특수한 대상에서 어떠한 현상이 유사하거나 일치하기 때문에 다른 현상도 유사하거나 일치할 것이라고 추론하는 방법
3. 생활 속의 알고리즘의 예
   1. 기초적인 수학의 연산
   2. 라면 조리법
   3. 하루의 일정 계획
   4. 가전제품의 사용 매뉴얼
   5. 지하철 환승 방법
   6. 알파고와 같은 인공지능 방법론
4. 전문가 시스템의 응용
   1. MYCIN
   2. DENDRAL
   3. PROSPECTOR
   4. AIRPLAN
5. 인공지능과 관련된 흥미로운 문제들
   1. 상자 쌓기
   2. 8-puzzle 게임
   3. 물통 문제
   4. 문자 암호 풀이
   5. 늑대, 염소, 양배추 문제

연습문제

진위 문제

1. 규칙기반 인공지능은 신경망과는 담당하는 역할이 다소 다르다. (O)

2. 규칙기반 인공지능에서는 논리를 바탕으로 규칙을 통해 추론한다. (O)

3. 논리의 중요한 목적에는 논리를 통한 입증이 옳은가를 포함하지 않는다. (X)

4. 논리는 일반적으로 명제 논리와 술어 논리로 구분된다. (O)

5. 논리에서 “3\*2=5라면, 런던은 미국의 수도다.”는 F가 된다. (X) – 조건 명제에서는 가정이 F이면 결론은 항상 T

6. 연역 추론은 하나 이상의 전제로부터 다른 어떤 결론을 도출해내는 추론 방식을 말한다. (O)

7. 역방향 추론은 출발 상태로부터 목표 상태를 향해 나가는 기법이다. (X)

8. 게임에서 나의 가능성을 최대로, 상대방의 가능성은 최소로 하는 전략 탐색을 쓴다. (O)

9. 세탁기나 전자레인지 등에서의 단계별 사용 설명서는 알고리즘에 속하지 않는다. (X)

10. 규칙이란 통상 “If ooo then xxx”이라는 형식으로 표현된다. (O)

단답식/선택식 문제

1. 논리와 명제는 (규칙기반) 인공지능에 필수적인 이론적 기반을 제공하고 있다.

2. (술어) 논리는 주어와 술어로 구분하여 참 또는 거짓에 관한 법칙을 다룬다.

3. 명확하게 참이나 거짓을 구분할 수 있는 문장이나 수학적 식을 (명제)라고 한다.

4. 추론은 크게 3가지로 나누면 귀납 추론, (연역) 추론, 그리고 유비 추론이다.

5. (귀납 추론)은 개별적 사실들의 분석으로부터 일반적인 결론을 끌어내는 추론 방법이다.

6. 목표를 찾기 위한 추론의 방법으로는 순방향 추론과 (역방향) 추론이 있다.

7. 문제 해결을 위한 여러 가지 단계들을 체계적으로 명시해 놓은 것이 (알고리즘)이다.

8. (전문가 시스템)은 인간 전문가들의 전문 지식을 수집 정리하여 인공지능의 추론 능력을 이용하여 해결하고자 하는 컴퓨터 자문 시스템의 일종이다.

9. 다음의 탐색 중 경험적 지식을 사용하는 탐색 방법은? (1)

(1) 휴리스틱 탐색 (2) 깊이 우선 탐색

(3) 너비 우선 탐색 (4) 무정보 탐색

10. 다음 중 규칙을 이용한 인공지능 기법을 통한 적절한 문제가 아닌 것은? (2)

(1) 물통 문제 (2) 복잡한 곱셈 문제

(3) 문자 암호 풀이 (4) 8-puzzle 게임

주관식 문제

1. 생활 속의 알고리즘의 예를 아는 대로 드시오.

- 기초적인 수학의 연산, 라면 조리법, 하루의 일정 계획, 가전제품의 사용 매뉴얼, 지하철 환승 방법, 알파고와 같은 인공지능 방법론

2. 연역 추론 중 삼단논법을 간단히 설명하시오.

- 미리 알려진 두 판단으로부터 그들과는 다른 하나의 새로운 판단으로 이끄는 추론 방법

3. 휴리스틱에 관해 간단하게 설명하시오.

- 어떤 사안 또는 상황에 대해 엄밀한 분석에 의하기보다 비슷한 문제에 대한 과거의 경험들을 바탕으로 직관적으로 판단하여 선택하는 의사결정 방식

4. 인공지능에서 문제 해결을 위한 주요 단계들을 기술하시오.

- 문제를 명확하게 정의해야 한다.

- 문제에 대한 철저한 분석이 필요하다.

- 정해진 제약 조건이나 규칙이 있는 경우 규칙의 적용에 대한 검증이 필요하다.

- 최적의 기법을 선택한다.

- 결과가 나오면 문제 해결 과정에 문제점이 없는지를 분석하고 검토한다.

5. 자동차 전문가 시스템이 할 수 있는 역할을 말해보시오.

- 손쉽게 전문가의 지식을 이용할 수 있다.

- 손상이 있는 부품이 있는지 등

8장 머신러닝의 학습과 분류

개념

1. 머신러닝: ‘기계학습’이라고도 불리는 인공지능의 한 분야인데, 인간의 학습 능력과 같은 기능을 컴퓨터와 같은 기계를 통해 실현하려는 기법과 분야이다  
   경험을 통해 데이터를 기반으로 학습하고 예측하는 시스템으로, 이를 구현하기 위해 알고리즘을 연구하고 구축하는 기술이라고 할 수 있다  
   프로그래머가 명시적으로 코딩하지 않고 자기 개선과 예제를 통해 학습할 수 있는 시스템이다
2. 지도 학습: 입력과 이에 대응하는 미리 알려진 출력을 연관시키는 관계를 학습하는 방법, 입력과 출력 쌍이 데이터로 주어지는 경우 그들 사이의 대응 관계를 학습하게 된다
3. 비지도 학습: 출력 없이 또는 출력값을 알려주지 않고 주어진 입력만으로 스스로 모델을 구축하여 학습하는 방법, 입력만 있고 출력 즉 레이블이 없는 경우에 적용하며, 입력 사이의 규칙성 등을 스스로 찾아내는 것이 학습의 주요 목표
4. 강화 학습: 주어진 입력에 대응하는 행동을 위하는 시스템에 대해 보상이 주어지게 되며, 이러한 보상을 이용하여 학습하는 방법, 주어진 입력에 대한 출력, 즉 정답 행동이 주어지지 않는다. 어떤 환경에서 정의된 대리자인 에이전트가 어떤 상태에서 시행착오를 통해 보상을 최대화하는 행동이나 행동 순서를 선택하는 머신러닝의 한 영역
5. 분류: 주어진 많은 데이터 중에 유사한 특성을 가진 것들끼리 묶어서 나누는 것
6. 회귀: 하나의 종속 변수와 다른 독립 변수들 사이의 관계를 결정하는 통계적 측정
7. 분류와 회귀의 차이점: 분류는 일정한 기준에 따라 명백하기 구분 짓는 것이고, 회귀는 오차 제곱의 합을 최소화하는 직선을 긋는 작업이다
8. 클러스터: 어떤 데이터 집합을 유사성에 따라 여러 개의 클래스나 그룹으로 나눌 때, 나누어진 클래스들
9. 클러스터링: 이처럼 나누는 작업, 서로 유사한 특성을 가진 여러 개의 그룹으로 묶는 기법인데, 같은 클러스터에 있는 점들은 다른 클러스터에 있는 점들보다 더 유사해야 한다
10. 지도 학습과 비지도 학습의 특징 비교(347p)

용어

1. 아서 새무엘: 머신러닝을 “프로그램을 명시적으로 작성하지 않고 컴퓨터에 학습할 수 있는 능력을 부여하기 위한 연구 분야이다.”라고 최초로 정의
2. 톰 미첼: “컴퓨터 프로그램이 어떤 작업 T와 평가 척도 P에 대해 경험 E로부터 학습한다는 것은, P에 의해 평가되는 작업 T의 성능이 경험 E에 의해 개선되는 경우를 말한다.”라며 더욱 구체적으로 정의
3. 전통적인 프로그래밍: 프로그래머가 관련 업무의 전문가와 상의하여 논리적 기반에 기초하여 모든 규칙들을 작성, 데이터와 규칙이 결합하여 출력을 생성
4. 머신러닝: 입력과 출력 데이터의 상관관계를 학습하여 규칙을 생성, 데이터와 출력이 함께 들어가서 규칙을 생성
5. 데이터 과학자: 머신러닝 기술을 활용하여 빅데이터를 분석
6. 이항 분류: 2개의 그룹으로 분류
7. 다항 분류: 3개 이상의 그룹으로 분류
8. 선형 회귀: 종속 변수 Y의 결과를 설명하거나 예측하기 위해 하나의 독립 변수를 사용하므로 비교적 간단
9. 회귀 직선: 직선 y = a + bx를 x에 대한 y의 회귀직선이라 한다
10. 베이즈의 정리: 과거의 데이터들을 기반으로 미래를 예측하는 모델로서 통계학과 경제학에 널리 적용되고 있으며, 검색 엔진, 스팸 메일 차단, 금융 이론, 승부 예측, 기상 예측, 의료 분야, 인공지능 등에 폭넓게 활용
11. 베이즈: 영국의 수학자이자 통계학자로서 확률에 대한 연구로 유명
12. 베이지안 네트워크: 방향성 비순환 그래프를 통하여 그 집합을 조건부 독립으로 표현하는 확률의 그래픽 모델, 추론과 학습을 수행하기 위한 효과적인 알고리즘이 존재한다
13. 은닉 마르코프 모델: 마르코프 모델의 일종으로, 시스템이 은닉된 상태와 관찰 가능한 결과의 두 가지 요소로 이루어진 확률형 모델이다

이해

1. 머신러닝의 종류
   1. 신경망: 생물의 신경 네트워크 구조와 기능을 모방한 모델
   2. 클러스터링: 주어진 데이터를 클러스터라는 부분 집합들로 분리하는 것
   3. 분류: 주어진 데이터를 비슷한 것들끼리 분류하는 것
   4. 의사결정 트리: 트리 구조 형태의 예측 모델로 의사를 결정하는 모델
   5. 나이브 베이즈: 베이즈 정리를 바탕으로 한 조건부 확률 모델 분류
2. 머신러닝의 활용 분야
   1. 영상인식: 문자인식, 물체인식
   2. 얼굴인식: Facebook에서의 얼굴인식
   3. 음성인식: Bixby, Siri, Alexa 등
   4. 자연어 처리: 자동 번역, 대화 분석
   5. 정보 검색: 스팸 메일 필터링
   6. 검색 엔진: 개인 맞춤식 추천 시스템
   7. 로보틱스: 자율주행 자동차, 경로 탐색
   8. 개인 비서, 헬스케어, 소셜 미디어 서비스, 기상 예측
3. 지도 학습의 장단점
   1. 장점  
      - 지도 학습을 사용하면 이전의 경험으로부터 데이터를 수집하거나 데이터 출력을 생성할 수 있다  
      - 지도 학습은 경험을 사용하여 성능 기준을 최적화할 수 있으며, 다양한 유형의 문제를 해결하는 데 도움이 된다
   2. 단점  
      - 지도 학습을 하기 위해서는 반드시 출력에 레이블이 있는 데이터들을 사용해야 하는 제한점이 있다  
      - 지도 학습에는 일반적으로 많은 시간이 걸리는데, 특히 데이터의 경우에는 엄청난 시간이 걸릴 수도 있다
4. 분류의 응용
   1. 사진으로 남자와 여자의 구별
   2. 개와 고양이의 구분
   3. 스팸 메일과 정상 메일 구분
   4. 0에서 9까지의 숫자의 구분
   5. 알파벳과 한글 문자 등의 구분
   6. 편지봉투의 손으로 쓴 주소 판별
   7. 카드 부정 사용 감지
   8. 의료 영상에서 종양의 존재 여부 판단
5. 회귀와 회귀 분석의 응용
   1. 날씨에 대한 예측
   2. 월별 판매액을 보고 다음 달 판매액 예측
   3. 금융, 투자, 비즈니스적 가격 판단
   4. 금값이나 원유 가격 예측
   5. 주택 가격의 예측
   6. 장단기 주가 예측
   7. 원유 가격 추정 등
6. 분류의 방법
   1. Naïve Bayes 분류기  
      머신러닝의 한 분야로서 자료가 여러 가지 속성을 가지고 있을 때, 해당 자료를 어느 클래스에 넣어야 할지를 베이즈 정리를 활용하여 판단해주는 알고리즘  
      나이브 베이즈 분류의 장점은 구축하기가 쉽고 대규모 데이터 세트에 매우 유용하다는 것이다
   2. 의사결정 트리  
      어떤 항목에 대한 관측값과 목표값을 연결하는 예측 모델인데, 통상 이진 트리를 사용한다  
      이진트리란 마치 나무를 거꾸로 세워놓은 모양인데, 최대 2가지의 판단을 하는 나무 모양의 구조이다
   3. SVM  
      머신러닝의 한 분야로서 배프닉이 1990년에 개발한 통계 학습 이론의 결과를 기반으로 한다  
      주어진 데이터를 2개의 영역으로 분류하는 이진 분류기인데, 기존 데이터를 바탕으로 새로운 데이터가 어느 영역에 속하는가를 판단하는 알고리즘
   4. K-Nearest Neighbor(K-NN)  
      1950년대에 개발된 지도 학습 모델의 분류 기법으로서 ‘최근접 이웃 분류’라고도 불리는데, 각 학습 데이터는 자기가 속하는 레이블을 가지는 분류기의 역할을 한다  
      새로 들어오는 입력 데이터 주위에 있는 k개의 가장 가까운 이웃 데이터를 선택하여, 이들 이웃 데이터들의 클래스 중 다수결로 하여 입력 데이터의 클래스를 결정하는 알고리즘
7. K-NN의 응용
   1. 영화나 음악 추천에 대한 개인별 선호 예측
   2. 수표에 적힌 광학 숫자와 글자인식
   3. 얼굴인식과 같은 컴퓨터 비전
   4. 유방암 등 질병의 진단
   5. 유전자 데이터의 인식
   6. 재정적인 위험성의 파악과 관리
   7. 주식 시장의 예측
8. 비지도 학습의 주요 응용 분야
   1. 비슷한 성향의 고객을 그룹으로 묶기
   2. 블로그에서 주제별로 구분하기
   3. 유사한 꽃이나 동물들끼리 묶기
   4. 네트워크상에서의 비정상적인 접근의 탐지
9. 주요 비지도 학습 방법
   1. K-means 클러스터링  
      주어진 데이터를 알고리즘에 따라 서로 유사한 특성을 가진 k개의 그룹으로 묶는 클러스터링 방법
   2. 가우스 혼합 모델
   3. 계층적 클러스터링
   4. 추천 시스템 등  
      추천을 위해 연관 데이터들을 정의하는 데 도움을 주는 클러스터링 방법
10. K-means 클러스터링은 다음과 같은 다양한 분야에 널리 활용되고 있다
    1. 통계: 주어진 데이터의 분류나 성향 분석
    2. 전자상거래: 고객의 구매 이력에 따른 고객 분류
    3. 건강 관리: 질병의 패턴이나 치료를 위한 패턴의 탐지
    4. 패턴: 이미지 검색을 위해 유사한 이미지를 그룹화
    5. 재무: 비정상적인 구매 패턴에 기반한 신용카드 사기 탐지
    6. 회사: 매출 등을 토대로 회사의 등급 분류
    7. 기술: 네트워크 침입 탐지나 악의적 활동의 탐지 시스템 구축
    8. 기상 예보: 폭풍의 예측을 위한 비정상적인 센서 데이터 탐지
11. 강화 학습의 응용 분야
    1. 통신망
    2. 로봇 제어
    3. 엘리베이터 제어
    4. 체스와 바둑 같은 게임

연습문제

진위 문제

1. 머신러닝은 인간의 학습 능력과 같은 기능을 실현하려는 기법과 분야이다. (O)

2. 머신러닝을 흔히 컴퓨터 프로그램이 데이터로부터 학습하는 과정이라고 말한다. (O)

3. 지도 학습을 통해서는 자동차 번호판의 인식률을 높일 수 없다. (X)

4. 비지도 학습에서는 데이터 학습을 위해 레이블이나 태그 표식을 붙인다. (X)

5. 회귀 분석 기술은 학습 데이터를 사용하여 하나의 출력값을 예측하는데, 산출물은 항상 확률론적 의미를 내포하고 있다. (O)

6. 분류는 일정한 기준에 따라 명백하게 구분 짓는 것이고, 회귀는 오차 제곱의 합을 최소화하는 직선을 긋는 작업이다. (O)

7. 머신러닝에서의 학습 방법 중 강화 학습은 포함되지 않는다. (X)

8. 머신러닝은 건강과 관련된 헬스케어 분야에서도 중요한 역할을 담당한다. (O)

9. 의사결정 트리는 이진 트리를 사용하는 관측값과 목표값을 연결하는 예측 모델이다. (O)

10. 강화 학습은 보상이 주어지는 문제 해결에 효과적인 학습 방법이 아니다. (X)

단답식/선택식 문제

1. (지도 학습)은 입력과 이에 대응하는 미리 알려진 출력을 연관시키는 관계를 학습하는 방법이다.

2. (비지도) 학습은 출력 없이 주어진 입력만으로 스스로 모델을 구축하여 학습하는 방법이다.

3. 스타크래프트와 바둑과 같은 게임에서는 보상을 위해 대부분 (강화) 학습이 사용된다.

4. (분류)란 주어진 데이터 중 유사한 특성을 가진 것들끼리 묶어서 나누는 것을 말한다.

5. (SVM) 알고리즘은 데이터를 경계선으로 구분하는데 그중 가장 큰 폭을 가진 경계선을 찾는 알고리즘이다.

6. (클러스터링)이란 서로 유사한 특성을 가진 여러 개의 그룹으로 묶는 기법이다.

7. (비지도) 학습이란 주어진 입력에 대응하는 출력 정보 없이 학습하는 것을 말한다.

8. (베이즈)의 정리는 과거의 데이터들을 기반으로 미래를 예측하는 모델로 활용된다.

9. 다음 중 머신러닝과 별로 관계가 없는 것은? (1)

(1) 프로그램을 명시적으로만 작성 (2) 전통적인 프로그래밍 개념과는 다름

(3) 인공지능에 속하는 부분 집합 (4) 데이터로부터 학습하여 지식을 획득

10. 다음 중 지도 학습의 장점에 속하는 것은? (2)

(1) 레이블이 있는 데이터 사용해야 함 (2) 경험을 사용하여 성능 기준을 최적화

(3) 일반적으로 많은 시간이 걸림 (4) 제한된 유형의 문제를 해결

주관식 문제

1. 머신러닝의 주요 종류들을 말하시오.

- 신경망: 생물의 신경 네트워크 구조와 기능을 모방한 모델

- 클러스터링: 주어진 데이터를 클러스터라는 부분 집합들로 분리하는 것

- 분류: 주어진 데이터를 비슷한 것들끼리 분류하는 것

- 의사결정 트리: 트리 구조 형태의 예측 모델로 의사를 결정하는 모델

- 나이브 베이즈: 베이즈 정리를 바탕으로 한 조건부 확률 모델 분류

2. 분류의 응용 예는 매우 다양한데 그중 몇 가지를 적으시오.

- 사진으로 남자와 여자의 구별

- 개와 고양이의 구분

- 스팸 메일과 정상 메일 구분

- 0에서 9까지의 숫자의 구분

- 알파벳과 한글 문자 등의 구분

- 편지봉투의 손으로 쓴 주소 판별

- 카드 부정 사용 감지

- 의료 영상에서 종양의 존재 여부 판단

3. 회귀의 개념을 설명하시오.

- 하나의 종속 변수와 다른 독립 변수들 사이의 관계를 결정하는 통계적 측정

4. 머신러닝에서 활용되는 비지도 학습 방법 몇 가지를 적으시오.

- K-means 클러스터링

- 가우스 혼합 모델

- 계층적 클러스터링

- 추천 시스템

5. 강화 학습에 대해 간단히 설명하시오.

- 주어진 입력에 대응하는 행동을 취하는 시스템에 대해 보상이 주어지며, 시행착오를 통해 보상을 최대화하는 행동이나 행동 순서를 선택하는 머신러닝의 한 영역

9장 신경망 개념과 응용

개념

1. 신경망: 인간 두뇌의 생물학적 뉴런의 작용을 모방하여 여러 뉴런들로부터 들어오는 입력을 일정한 함수를 거쳐 출력 노드를 통해 결과를 얻는 네트워크로서, 인공신경망으로 부르기도 한다
2. 신경망 모델의 시초: 1943년 워렌 맥컬럭과 월터 피츠의 연구에서 비롯되었다. 그들은 인간의 두뇌를 논리적 서술을 구현하는 ON이나 OFF의 이진 원소들의 집합으로 추측하였으며, 수많은 뉴런들로 이루어진 잘 정의된 컴퓨터라고 여겼다
3. 신경망에서 뉴런에 해당하는 노드는 비선형적이다

용어

1. 도널드 헵: 1949년 제안한 ‘헵의 학습 규칙’은 두 뉴런 사이의 연결강도를 조절할 수 있는 최초의 규칙
2. 로젠블럿의 단층 퍼셉트론: 최초의 신경망 모델
3. 마크 I 퍼셉트론: 퍼셉트론 이론에 의해 만들어진 최초의 신경망 하드웨어 장치
4. 선형 분리 가능: 패턴 클래스가 하나의 직선에 의해 두 개의 영역으로 나누어지는 것
5. 아달라인: 위드로와 호프가 제안한 알고리즘을 사용하여 퍼셉트론과 비슷한 시기에 나타난 신경망 모델, 이것을 여러 개 묶은 마달라인도 있다
6. 다층 퍼셉트론: 입력층과 출력층 사이에 하나 이상의 은닉층이 존재하는 신경망으로 계층 구조를 가진다
7. 역전파 알고리즘: 입력층으로부터 은닉층을 거쳐 출력층으로 갔다가, 다시 반대 방향으로 되돌아오면서 학습
8. 델타규칙: 입출력 함수가 선형의 노드로부터 이루어진 네트워크에 대하여, 모든 입력 패턴으로부터 얻어지는 출력과 목표 출력과의 오차 제곱의 총합을 최소화하도록 연결강도를 조정하는 규칙
9. 최급하강법: 곡면에 대하여 오차의 제곱이 가장 많이 감소하는 방향으로 기울기를 따라가며 변화하는 방법

이해

1. 신경망의 응용 분야
   1. 문자인식
   2. 음성인식
   3. 영상인식
   4. 자연어 처리
2. 지금까지의 신경망 모델과 이에 대응하는 알고리즘
   1. 1957년 로젠블럿의 퍼셉트론 모델(퍼셉트론 알고리즘)
   2. 1984년 PDP 그룹 이후의 다층 퍼셉트론 모델(역전파 알고리즘)
   3. 2006년 힌턴 이후의 심층신경망(딥러닝 알고리즘)
3. 신경망에서 많이 사용되는 대표적인 활성 함수
   1. 계단 함수
   2. 임계논리 함수
   3. 시그모이드 함수
4. 퍼셉트론의 학습 과정
   1. 연결강도들과 임계값을 초기화한다
   2. 새로운 입력과 기대되는 출력을 제시한다
   3. 실제 출력값을 계산한다
   4. 연결강도를 재조정한다
   5. 더 이상 조정이 없을 때까지 2단계로 가서 반복 수행한다
5. 단층 퍼셉트론의 2가지 제한점
   1. 퍼셉트론의 출력은 0 또는 1만 가진다는 제한점이 있다
   2. 퍼셉트론은 오직 선형 분리가 가능한 집합만을 분리할 수 있다는 점이다
6. 아달라인의 응용 분야
   1. 시스템 모델링
   2. 통계적 예측
   3. 통신 잡음과 울림 제거
   4. 채널 이퀄라이저
   5. 적응적 신호 처리

연습문제

진위 문제

1. 뉴런의 작용을 모델링한 신경망은 인공신경망으로도 불린다. (O)

2. 맥컬럭은 인간 두뇌를 수많은 뉴런들로 이루어진 잘 정의된 컴퓨터라고 여겼다. (O)

3. 신경망은 문자인식, 음성인식, 영상인식, 자연어 처리 등의 분야에 이용되고 있다. (O)

4. XOR 함수는 선형 분리가 가능한 논리함수에 속한다. (X)

5. 신경망에서 가장 간단한 노드는 n개의 입력을 받아 n개의 연결강도 벡터들과 각각 곱해진 결과가 합해져서 특정한 활성 함수를 거쳐 출력을 낸다. (O)

6. 노드는 내부적인 임계값이나 오프셋, 그리고 비선형 함수의 형태에 따라 그 값이 정해지게 된다. (O)

7. 단층 퍼셉트론의 한계점이 노출되면서 2000년대 중반에 다층 퍼셉트론 모델이 제안되었다. (X)

8. 신경망은 병렬처리나 학습과 관련된 지능적인 역할을 훌륭하게 수행해낸다. (O)

9. 단층 퍼셉트론은 딥러닝의 심층신경망을 거쳐 다층 퍼셉트론으로 발전하였다. (X)

10. 신경망에서 계산의 복잡성으로 학습 시간이 너무 오래 걸리는 등의 문제점이 있다. (O)

단답식/선택식 문제

1. 신경망은 인간 두뇌의 생물학적 (뉴런)의 작용을 모방하여 고안되었다.

2. 신경망에서는 뉴런 사이의 (연결강도) 조정을 통해 학습이 가능하다.

3. 다층 퍼셉트론의 작동은 왕복 운동을 하는 (역전파) 학습 알고리즘에 이루어진다.

4. 다층 퍼셉트론은 입력층, (은닉층), 출력층의 순서와 방향으로 연결되어 있다.

5. 민스키와 페퍼트는 (퍼셉트론즈)란 저서에서 단층 퍼셉트론의 문제점들을 밝혀냈다.

6. 마크 I 퍼셉트론은 (문자)를 인식하는 놀라운 성과에 많은 사람들이 환호했다.

7. (델타규칙)이란 모든 입력 패턴으로부터 얻어지는 출력과 목표 출력과의 오차 제곱의 총합을 최소로 하도록 연결강도를 조정하는 규칙이다.

8. 다층 퍼셉트론에서는 매우 낮은 확률이지만 (지역) 최소점 문제에 봉착할 수 있다.

9. 다음 중 신경망에서 많이 사용되는 비선형 함수가 아닌 것은? (4)

(1) 계단 함수 (2) 시그모이드 함수

(3) 임계논리 함수 (4) 사인 함수

10. 다음 중 선형 분리가 불가능한 논리함수는? (2)

(1) AND 함수 (2) XOR 함수

(3) OR 함수 (4) NOT 함수

주관식 문제

1. 신경망의 획기적인 3가지 모델과 그에 해당하는 알고리즘을 적으시오.

- 1957년 로젠블럿의 퍼셉트론 모델(퍼셉트론 알고리즘)

- 1984년 PDP 그룹 이후의 다층 퍼셉트론 모델(역전파 알고리즘)

- 2006년 힌턴 이후의 심층신경망(딥러닝 알고리즘)

2. 최초의 학습 규칙은 무엇이고 개발한 인물은 누구인가?

- 헵의 학습 규칙

- 도널드 헵

3. 선형 분리 가능과 특징을 설명하시오.

- 패턴 클래스가 하나의 직선에 의해 두 개의 영역으로 나누어지는 것을 말한다.

- 논리에서 AND와 OR 함수는 진리표의 좌표에 따라 하나의 직선에 의해 분리되는 선형 분리가 가능하다.

- 그러나 Exclusive-Or(XOR) 함수는 선형 분리가 가능하지 않은 패턴분류 문제의 대표적인 예이다.

4. 신경망이 문자인식에 많이 쓰이는 이유와 단점을 기술하시오.

- 신경망은 특성상 잡음이나 왜곡, 크기의 다양성, 위치 변화 등에 비교적 잘 적응할 수 있으므로 문자인식에 많이 쓰이고 있다.

- 그러나 신경망에 의한 문자인식에는 훈련하는 데 많은 시간이 걸리는 단점을 가지고 있다.

5. 다층 퍼셉트론에서 역전파 알고리즘을 설명하시오.

- 다층 퍼셉트론 모델에 사용되는 역전파 학습 알고리즘은 전방향과 역방향으로 반복적으로 움직이면서 역전파 학습을 하게 된다.

10장 딥러닝 기반의 심층신경망과 활용

개념

1. 딥러닝: 머신러닝과 신경망의 한 분야로서 여러 개의 은닉층을 가진 심층신경망을 기반으로 하는 학습 방법이다
2. 딥러닝의 두드러진 장점: 머신러닝과 달리 특징 추출과 분류가 자동적으로 이루어짐

용어

1. 네오코그니트론: 딥러닝의 시초로 1980년 초반 일본의 후쿠시마 박사가 개발
2. RBM: 새로운 딥러닝 기반의 학습 알고리즘
3. 컨볼루션 신경망: ‘합성곱 신경망’이라고도 하는데, 영상의 분석이나 영상인식에 많이 사용되는 심층신경망의 한 종류이다. 합성곱 연산을 사용하는데, 합성곱을 사용하면 3차원 데이터의 공간적 데이터를 유지한 채 다음 층으로 보낼 수 있다. 동물의 시신경 구조와 유사하게 뉴런 사이의 연결 패턴을 형성하는 모델로서 특징 지도를 이용하여 학습한다.
4. 순환 신경망: 순차적 정보가 담긴 데이터에서 규칙적인 패턴을 인식하고, 추상화된 정보를 추출할 수 있는 모델이다. 노드 간의 연결이 순환적 구조를 가지는 것이 특징인데, 필기체 인식이나 음성인식과 같이 시간에 따라 변하는 특징을 가지는 데이터를 잘 처리할 수 있다
5. 제한된 볼쯔만머신(RBM): 힌턴이 제안한 모델로, 비지도 학습에 활용된다. 입력 집합에 대한 확률 분포를 학습할 수 있는 신경망인데 확률은 에너지 함수 형태로 표현되며, 에너지가 최소화되는 방향으로 학습한다
6. 심층 신뢰 신경망: 다층의 잠재 변수로 표현하는 은닉층으로 이루어지는 심층신경망인데, 사전에 훈련된 RBM을 여러 층으로 쌓아 올린 구조이다
7. 생성적 적대 신경망: 2014년 이안 굿펠로우 등에 의해 발표되었는데, 차세대 딥러닝 알고리즘으로 주목받고 있다. 진짜 같은 가짜를 생성하는 모델과 이에 대한 진위를 판별하는 모델의 경쟁을 통해 진짜 같은 가짜 이미지를 만들어낼 수 있다
8. XAI(설명가능 인공지능): 딥러닝을 통한 인식 결과만 알려주는 인공지능을 넘어 그 이유까지 설명
9. NPU: 퀄컴과 삼성전자 등에서 개발한 딥러닝 학습 알고리즘을 병렬로 처리할 수 있는 하드웨어
10. 텐서플로: 구글의 브레인팀에서 개발한 다양한 작업에 대해 데이터플로 그래프를 사용하여 머신러닝과 딥러닝 프로그램을 쉽게 구현할 수 있도록 다양한 기능을 제공해주는 소프트웨어 라이브러리
11. 파이토치: 페이스북의 인공지능팀이 개발한 Python 기반의 오픈소스 머신러닝 라이브러리
12. CNTK: 마이크로소프트의 코그니티브 툴킷(CNTK)은 텐서플로처럼 그래프 구조를 사용하여 데이터 흐름을 기술하며 딥러닝 신경망을 구성
13. 아파치 MXNet: 아마존의 오픈소스 딥러닝 소프트웨어로서 아마존 웹 서비스의 고급 딥러닝 프레임워크로 채택하고 있다

이해

1. 신경망과 심층신경망의 비교
   1. 신경망 심층신경망  
      은닉층의 개수 통상 1~2개 1개 이상 1,000개까지  
      알고리즘 역전파 알고리즘 딥러닝 알고리즘  
      주요 이용 시대 1986년 이후 2006년 이후  
      사용하는 데이터 학습용 데이터 정도 대규모 데이터  
      이론상 수행 시간 실행 시간이 많이 걸림 은닉층 개수가 많아  
       실행 시간이 훨씬 많이 걸림  
      프로세서 속도 당시 상대적으로 느림 최근에 크게 빨라짐  
      병렬처리 병렬 슈퍼컴퓨터를 사용하였으나 수많은 고속 프로세서들을  
       큰 성과 없음 병렬로 연결하여 빠르게 사용 가능
2. 머신러닝과 딥러닝의 특징 비교
   1. 머신러닝 딥러닝  
      데이터 크기 작은 데이터 집합에 좋은 결과 매우 큰 데이터 집합에 좋은 결과  
      처리하는 컴퓨터 학습 시간이 짧아 일반 컴퓨터 매우 오래 걸리므로 강력한 컴퓨터  
      특징 추출의 방법 최상의 결과를 위해 여러 가지 특징 추출과 분류가 자동적으로 처리 가능  
       특징 추출과 분류 방법 시도  
      처리 시간 몇 분에서 몇 시간 정도 경우에 따라 몇 주까지도 걸림  
      알고리즘의 종류 다양하고 많음 현재는 적으나 많이 개발 중임
3. 딥러닝 심층신경망의 종류별 주요 용도
   1. 컨볼루션 신경망: 영상인식, 컴퓨터 비전
   2. 순환 신경망: 음성인식, 작곡, 주가 예측
   3. 제한적 볼쯔만머신 신경망: 분류, 회귀 분석
   4. 심층 신뢰 신경망: 글씨와 음성의 인식
   5. 생성적 적대 신경망: 영상과 음성의 복원
4. 딥러닝의 활용과 연구 동향
   1. 구글  
      음성인식과 번역을 비롯한 인공지능 시스템 개발에도 딥러닝 기술을 적용하고 있다
   2. 마이크로소프트  
      2014년 ‘아담 프로젝트’라는 딥러닝 기술을 공개했다
   3. 페이스북  
      2014년 딥러닝 기술을 적용하여 ‘딥페이스’라는 얼굴인식 알고리즘을 개발했다
   4. 트위터  
      ‘매드비츠’라는 딥러닝 회사를 인수하여 사진 분석 기술을 확보하고 있다
   5. 딥 마인드  
      딥 마인드 개발자인 하사비스를 중심으로 최근 새로운 딥러닝 기술을 제시하여 많은 관심을 끌고 있다
   6. 네이버  
      음성인식, 뉴스 요약, 이미지 분석 등에 딥러닝 기술을 적용하고 있다
5. 텐서플로의 활용
   1. 필기체 글씨나 숫자의 인식
   2. 음성인식
   3. 영상인식
   4. 자연어 처리
   5. 기계 번역
6. PyTorch의 장점
   1. 설치가 간편하며 이해와 디버깅이 쉬운 직관적이고 간결한 코드로 구성되었다
   2. 실시간으로 결과값을 시각화해준다
   3. Python 라이브러리(Numpy, Scipy, Cython)와 높은 호환성을 가진다
   4. 모델 그래프를 만들 때 데이터에 따라 유연한 조절이 가능하다
   5. 학습 및 추론 속도가 빠르고 다르기 쉽다

연습문제

진위 문제

1. 딥러닝 기반의 심층신경망은 이미 존재하는 다층 신경망의 특수한 경우로 볼 수 있다. (O)

2. 딥러닝의 시초는 1980년 일본의 후쿠시마가 개발한 네오코그니트론이다. (O)

3. 딥러닝은 머신러닝이나 신경망보다 훨씬 큰 규모의 데이터를 학습할 수 없다. (X)

4. 힌턴은 새로운 딥러닝 기반의 학습 알고리즘을 제안하면서 오랫동안 미해결된 과적합 문제 등을 해결할 수 있게 되었다. (O) – RBM을 제안하고, 연이어 드롭아웃이라는 알고리즘도 등장하여 과적합 문제 등을 해결

5. RBM이란 제한된 볼쯔만머신은 지도 학습에만 활용된다. (X)

6. 심층 신뢰 신경망은 사전에 훈련된 RBM을 여러 층으로 쌓아 올린 구조이다. (O)

7. 컴퓨터 게임의 지능적 행동은 딥러닝 학습을 통해 다양하게 적용되고 있다. (O)

8. 인공지능 반도체 시장은 지금부터는 다소 정체될 것으로 예측된다. (X)

9. 텐서플로는 누구나 무료로 다운받아 사용할 수 있는 소프트웨어가 아니다. (X)

10. 텐서플로 자체는 C++로 구현되어 있으며, Python, Java 등의 언어를 지원한다. (O)

단답식/선택식 문제

1. 2006년 힌턴 교수는 다층 신경망에다 학습을 통한 전처리 과정을 추가한 (딥러닝) 학습 알고리즘 기법을 발표하였다.

2. 심층신경망은 여러 개의 (은닉층)을 사용하였는데, 많을 경우에는 무려 1,000여 개가 사용되기도 한다.

3. 딥러닝의 장점으로는 (특징 추출)과 분류가 자동적으로 이루어진다는 점이다.

4. (순환) 신경망은 순차적 정보가 담긴 데이터에서 규칙적인 패턴을 인식하고, 추상화된 정보를 추출할 수 있는 모델이다.

5. (텐서)란 다양한 차원의 배열로 나타내는 데이터를 말한다.

6. (깃허브)는 개발자의 성지라고 불리며, 오픈소스 소프트웨어의 중심지 역할을 한다.

7. (텐서플로)는 딥러닝을 지원하는 구글이 만든 소프트웨어이다.

8. 텐서플로의 대부분의 편리한 기능들이 (Python) 라이브러리로 구현되어 있다.

9. (파이토치)는 페이스북이 개발한 Python 기반의 오픈소스 머신러닝 라이브러리이다.

10. 다음 중 텐서플로가 자주 활용되지 않는 것은? (3)

(1) 영상인식 (2) 음성인식

(3) 적분 계산 (4) 자연어 처리

주관식 문제

1. 딥러닝이 무엇인가와 응용 분야를 간단히 설명하시오.

- 딥러닝이란 머신러닝과 신경망의 한 분야로서 여러 개의 은닉층을 가진 심층신경망을 기반으로 하는 학습 방법이다.

- 일반적으로 딥러닝 학습은 영상이나 음성 등 대량의 데이터로부터 특징을 추출하여 학습을 통해 음성인식, 영상인식 등 패턴인식에 좋은 성과를 나타낸다.

2. 세계적인 IT 기업들의 딥러닝 연구 개발 상황을 간략하게 요약하시오.

- 구글은 음성인식과 번역을 비롯한 인공지능 시스템 개발에도 딥러닝 기술을 적용하고 있다.

- 마이크로소프트는 2014년 ‘아담 프로젝트’라는 딥러닝 기술을 공개했다.

- 페이스북은 2014년 딥러닝 기술을 적용하여 ‘딥페이스’라는 얼굴인식 알고리즘을 개발했다.

- 트위터는 ‘매드비츠’라는 딥러닝 회사를 인수하여 사진 분석 기술을 확보하고 있다.

- 딥 마인드는 최근 새로운 딥러닝 기술을 제시하여 많은 관심을 끌고 있다.

- 네이버는 음성인식, 뉴스 요약, 이미지 분석 등에 딥러닝 기술을 적용하고 있다.

- 다음카카오도 자회사를 통해 딥러닝 기술을 활용하고 있다.

3. 설명가능 인공지능의 핵심을 간략히 설명하시오.

- 설명가능 인공지능(XAI)란 딥러닝을 통한 인식 결과만 알려주는 인공지능을 넘어 그 이유까지 설명해준다.

4. NPU가 무엇인가를 간략하게 설명하시오.

- NPU는 인간의 뇌와 같은 역할을 할 수 있는 차세대 반도체로서, 여러 가지 복잡한 연산의 실시간 처리가 가능하다.

5. 최근 파이토치의 인기가 점차 높아지고 있는 이유를 말하시오.

- 파이토치의 인기가 높아지고 있는 이유로는 파이토치가 간결하고 구현이 빠르며, 사용자가 텐서플로보다 훨씬 익히기 쉽다는 점이다. 또 코드를 다른 사람들에게 설명해주기에도 매우 효과적이기 때문이다.

11장 인공지능의 패턴인식

개념

1. 패턴인식: 패턴이나 특징적인 경향을 발견하여 인식하는 것
2. 음성인식: 마이크와 같은 컴퓨터 입력 장치를 통해 사람의 음성을 입력 받은 후 음성에 포함된 정보를 추출하여 문자 데이터로 전환하는 처리
3. 영상인식: 디지털 형태의 영상 파일을 신경망이나 딥러닝 기술로 분석하여 인식하는 것
4. 자연어: 인간이 태어날 때부터 익혀 일상생활에서 사용하고 있는 언어

용어

1. 화자종속: 특정한 사람의 말만 인식하고 다른 사람의 말은 인식하지 못하는 것
2. 화자독립: 일반적인 발음을 하는 대다수 사람들의 음성을 인식할 수 있다
3. 음성합성: 음성인식과 반대로 문자로 된 정보를 사람이 알아들을 수 있는 음성으로 만들어주는 기술로서, TTS라고도 한다
4. TypeCast: 인공지능을 이용한 음성합성 서비스
5. 픽셀: 영상을 구성하는 최소 단위인 점
6. 영상이해: 주어진 영상을 보고 그 내용까지 이해할 수 있는 상당히 지능적인 단계
7. 형식 언어: 일정한 문법에 맞춰 만들어진 컴퓨터용 언어로서, Python이나 C와 같은 프로그래밍 언어가 이에 해당한다
8. 자연어 이해: 글을 이해하여 요약하거나 주어진 자연어를 통해 상황을 이해하는 수준
9. 기계번역: 컴퓨터를 통해 서로 다른 언어들을 번역하는 일
10. 구문론적 이해: 문법적인 지식을 토대로 함
11. 의미론적 이해: 문맥 정보나 상식에 따른 추론에 의함
12. 자연어 자동 번역 프로그램: 컴퓨터와 인공지능 기술을 이용하여 문맥과 의미를 파악하여 한글, 영어, 중국어, 아랍어 등 다양한 언어로 번역해주는 프로그램
13. 이미지 번역: 스마트폰 카메라로 외국어를 비추면 설정해 놓은 언어로 즉석에서 번역되는 기능
14. 파파고: 네이버 랩에서 자체 개발하여 무료로 제공 중인 국내 토종 인공지능 번역기

이해

1. 패턴인식의 응용
   1. 문자인식
   2. 음성인식
   3. 영상인식
   4. 지문인식
   5. 홍체인식
   6. 동영상에서의 물체인식
   7. 자연어 처리
2. 음성인식의 응용
   1. 음성에 의한 기계의 조작
   2. 음성타자기
   3. 로봇 제어
   4. 무인 전화 예약 시스템
3. 영상인식의 응용
   1. 자동차 번호판 인식
   2. 로봇의 영상인식
   3. 지문인식과 홍체인식
   4. 딥러닝을 이용한 얼굴인식 시스템
   5. 자율자동차에서의 물체인식
   6. 영상이해

연습문제

진위 문제

1. 음성인식은 발성자의 성별이나 나이, 지속 시간 등에 따라 인식 결과가 다르다. (O)

2. 영상인식은 2차원이나 3차원의 영상 정보를 다루므로 인식 시간이 상당히 빠른 편이다. (X)

3. 화자종속 음성인식 기술이 화자독립 음성인식 기술보다 훨씬 더 어려우며 실제 생활에 널리 활용될 수 있다. (X)

4. 음성합성 시스템은 주어진 문장을 음성으로 들을 수 있는 편리함을 제공해줄 수 있다. (O)

5. 대부분의 최신 스마트폰에는 음성인식을 통해 비서 역할을 하는 에이전트가 적용된다. (O)

6. 지문인식 기술은 보안이 필요한 장소에 출입자의 신분 확인에 많이 쓰이고 있다. (O)

7. 문자인식은 영상인식의 영역에 속하지 않는다. (X)

8. 자연어에서 문법적 지식에다 의미를 부여하는 것은 구문론적 이해이다. (O)

9. 자연어 이해란 글을 이해하여 자연어 표현의 상황을 이해하는 수준을 말한다. (O)

10. 네이버의 번역 서비스인 파파고는 10여 개 언어의 번역 서비스를 제공 중이다. (O)

단답식/선택식 문제

1. 키보드 없이 말하는 음성을 인식하여 마치 타자기를 치듯이 문서로 바꾸는 장치는?

- 음성타자기

2. 눈동자의 홍채 정보를 이용하여 사람을 인식할 수 있는 기술은 무엇인가?

- 홍체인식

3. 문자로 된 정보를 사람이 알아들을 수 있게 음성으로 만들어주는 기술은?

- 음성합성

4. 음성인식 기술은 화자의 대상에 따라 화자종속과 (화자독립)의 2가지로 나눌 수 있다.

5. (영상인식)이란 영상 파일을 신경망이나 딥러닝 기술로 분석하여 인식하는 기술이다.

6. (영상이해)란 주어진 영상을 보고 그 내용까지 이해할 수 있는 상당히 지능적인 단계이다. 이것은 영상인식과 지식처리 기술이 융합되어야만 가능하다.

7. 인공지능 목소리는 (음성합성)과 인공지능 기술의 결합으로 이루어질 수 있다.

8. 컴퓨터를 통해 서로 다른 언어들을 번역하는 일을 (기계번역)이라 한다.

9. 다음의 인식 중 접촉식 인식방법에 해당되는 것은? (2)

(1) 홍체인식 (2) 지문인식

(3) 얼굴인식 (4) 자동차 번호판 인식

10. 다음 중 현재의 패턴인식 기술과 관계가 비교적 적은 것은? (3)

(1) 문자인식 (2) 음성인식

(3) 감정인식 (4) 얼굴인식

주관식 문제

1. 패턴인식에서 응용되는 대표적인 종류를 몇 가지 적으시오.

- 로봇의 시각, 청각

- 우편번호 자동인식

- 지문 자동판별

2. 음성인식을 위해 많이 쓰이는 몇 가지 방법을 말하시오.

- 특징 추출을 기반으로 하는 패턴 매칭을 이용하는 방법

- 확률적인 성질을 기반으로 하는 은닉 마르코프 방법

- 자연어 처리에서의 문맥을 고려한 방법

- 신경망에 의한 방법

- 딥러닝에 의한 방법

3. 출입자의 신분을 확인하는 데 많이 쓰이고 있는 3가지 인식 방법을 말하시오.

- 지문인식

- 홍체인식

- 얼굴인식

4. 형식 언어의 정의와 특징을 간단히 설명하시오.

- 일정한 문법에 맞춰 만들어진 컴퓨터용 언어로서, Python이나 C와 같은 프로그래밍 언어가 이에 해당한다.

- 콤마 하나만 빠져도 실행될 수 없으며, 문법에 따라 매우 정확해야 하며, 애매한 점이 전혀 없어야 한다.

5. 자연어 처리에서 구문론과 의미론의 차이점을 설명하시오.

- 구문론적 이해는 문법적인 지식에 토대를 둔 방법이다.

- 의미론적 이해는 문맥 정보나 상식에 따른 추론에 의한 방법이다.

12장 인공지능과 데이터 사이언스

개념

1. 데이터 사이언스: 통계 방법을 사용하여 대량의 데이터를 처리하는 여러 학문 분야가 관련된 과학으로서, 데이터에 대한 직관력을 획득할 수 있게 해주는 학문 분야
2. 빅데이터: 데이터는 ‘4차 산업혁명의 쌀’이라고 불리는데, 다양한 데이터를 많이 모아 놓은 것이 바로 빅데이터이다
3. 빅데이터 기술: 엄청난 양의 정보를 빠르게 분석하여 의미 있는 결론을 도출해내는 것
4. 데이터 마이닝: 대규모 데이터 집합으로부터 통계적 규칙이나 패턴을 체계적으로 찾아내는 과정

용어

1. 데이터 과학자: 데이터 사이언스와 관련된 업무의 전문가
2. 시각화: 분석의 결과를 눈에 잘 띄도록 그래프나 표로 표현하는 것
3. 분류: 주어진 많은 데이터 중에서 비슷한 특성을 가진 것들끼리 묶어서 나누는 것
4. 클러스터링: 데이터의 집합을 유사도에 따라 비슷한 클래스나 클러스터들로 나누는 것

이해

1. 데이터 사이언스가 큰 관심을 끌면서 여러 분야에 폭넓게 활용되고 있는 이유
   1. 첫째, 처리 속도가 빠르고 널리 보급되어 활용되고 있는 컴퓨터 기술
   2. 둘째, 인터넷을 통해 손쉽게 접할 수 있는 풍부하고 다양한 데이터
   3. 셋째, 인공지능에서 머신러닝을 이용하여 빠르고 편리하게 구현
2. 데이터 사이언스 응용 분야
   1. 경영
   2. 인공지능
   3. 사회학
   4. 의학
   5. 생물학
   6. 인문학
3. 데이터 사이언스의 비즈니스에의 접목
   1. 의사 결정의 향상
   2. 경영상의 문제 개선
   3. 미래에 대한 예측
   4. 경영 목표 개선
   5. 새로운 기회 창출
   6. 위험 평가 향상
   7. 데이터 기반 습관
4. 데이터 사이언스에 필요한 지식과 기술
   1. 컴퓨터 지식
   2. 프로그래밍 기술
   3. 통계처리의 기초 지식
   4. 머신러닝과 관련된 지식과 기술
   5. 데이터를 분석한 결과를 시각화
5. 데이터 사이언스 도구
   1. Python
   2. R
   3. SAS
   4. Apache Spark
   5. TensorFlow
6. 데이터 사이언스 작업의 흐름
   1. 문제 정의  
      명확한 문제 정의와 반드시 이루어져야 할 목표값을 설정한다
   2. 데이터 수집  
      필요한 데이터들을 데이터베이스나 웹 사이트 등으로부터 수집한디
   3. 데이터 관리  
      수집된 데이터를 적절하게 저장하고 관리한다
   4. 탐색적 분석  
      데이터로부터 패턴을 찾아내고 탐색적 분석을 수행한다
   5. 최종 분석  
      탐색적 분석 단계에서 수행했던 과정을 다양한 항목의 체크 리스트를 이용하여 최종적으로 점검한다
   6. 보고  
      최종 분석의 결과를 바탕으로 보고서를 작성하여 제출한다
7. 빅데이터의 특징
   1. 크기
   2. 변화의 다양성
   3. 속도
   4. 진실성
   5. 시각화
   6. 가치
8. 빅데이터의 요소 기술
   1. 빅데이터 수집
   2. 빅데이터 공유
   3. 빅데이터 저장
   4. 빅데이터 처리
   5. 빅데이터 분석
   6. 빅데이터 시각화
9. 빅데이터의 활용 사례
   1. 한국석유공사의 국내 유가 예측 서비스
   2. 구글의 검색어 분석을 통한 독감 예보 서비스
   3. 싱가포르의 스마트시티 구현을 위한 프로젝트
   4. 영국 런던의 교통국에서는 교통, 환경, 공해 문제를 해결하기 위해 빅데이터를 적극적으로 활용
   5. 주차장에서의 차량인식
   6. 버스 노선의 실선과 배차 간격
10. 데이터 마이닝 업무의 중점
    1. 데이터로부터 정보를 추출하는 일
    2. 감춰진 패턴들을 발견해 내는 일
    3. 예측 모델을 개발하는 일
11. 데이터 마이닝의 기능
    1. 분류 기능
    2. 클러스터링 기능
    3. 연관성 기능
    4. 연속성 기능
    5. 예측 기능
12. 데이터 마이닝의 활용
    1. 기업의 생산 과정에서 불량률을 줄이는 품질관리 분야
    2. 패턴인식 기법을 적용한 의료 진단 분야
    3. 고객의 신용을 평가하는 금융 관리 분야
13. 데이터 사이언스의 미래
    1. 데이터 사이언스의 역할 증대
    2. 데이터 사이언스의 영역이 명확히 정의됨
    3. 관련 직업이 많아짐
    4. 데이터 사이언스 교육이 일반화됨
    5. 데이터 사이언스를 위한 머신러닝의 발전

연습문제

진위 문제

1. 데이터 사이언스는 통계 방법만을 사용하여 대량의 데이터를 처리하는 전문 분야다. (X)

2. 데이터 사이언스는 데이터를 다루는 방법론, 프로세스, 알고리즘 등을 다루게 된다. (O)

3. 머신러닝을 전혀 이용하지 않고도 데이터 사이언스가 폭넓게 활용될 수 있다. (X)

4. SAS는 편리한 연산을 위해 통계적 소프트웨어로서 안정성과 신뢰성을 가지고 있다. (O)

5. 데이터 사이언스에서의 분석은 분류, 회귀, 그리고 클러스터링에 의한 경우가 많다. (O)

6. 빅데이터는 구매 패턴 분석과 앱에서의 경향 분석 등에 응용될 수 있다. (O)

7. 태풍이나 허리케인과 관련된 빅데이터를 활용하여 재해 예방에 활용하고 있다. (O)

8. 개인정보보호 문제는 빅데이터의 자료 수집 문제와 거의 관련이 없다. (X)

9. 데이터 마이닝은 데이터로부터 정보를 추출하는 일, 패턴들을 발견해 내는 일, 그리고 예측 모델을 개발하는 등에 중점을 둔다. (O)

10. 우리나라에서는 아직까지 데이터 거래소가 별로 필요하지 않다. (X)

단답식/선택식 문제

1. 데이터 사이언스나 빅데이터와 관련된 업무의 전문가를 (데이터 과학자)라 한다.

2. 다양한 형태의 정형 또는 비정형 데이터를 많이 모아 놓은 것이 (빅데이터)이다.

3. 빅데이터의 3가지 특징(V3)은 크기, 변화의 다양성, (속도)이다.

4. (데이터 마이닝)은 대규모 데이터 집합으로부터 통계적 규칙이나 패턴을 체계적으로 찾아내거나 캐내는 뜻을 가진 과정이다.

5. (분류)란 주어진 많은 데이터 중 비슷한 특성을 가진 것들끼리 분리하는 것이다.

6. (클러스터링)이란 데이터의 집합을 유사도에 따라 클러스터들로 나누는 것이다.

7. (시각화)란 탐색적 분석의 결과가 나오면 그것을 보기 쉽고 잘 이해할 수 있도록 하는 것이다.

8. (R)은 오픈소스 통계적 프로그래밍 언어인데, 데이터를 분석하고 가시화하는 데 도움을 줄 수 있는 다양한 패키지를 제공한다.

9. 데이터 사이언스에 필요한 지식과 기술 중 중요성이 상대적으로 적은 것은? (2)

(1) 데이터 수집 기술 (2) 컴퓨터 하드웨어 지식

(3) 통계처리와 시각화 기술 (4) 머신러닝과 관련된 지식과 기술

10. 다음 중 데이터 사이언스의 도구 중 가장 관계가 먼 것은? (4)

(1) Python (2) R

(3) SAS (4) Ruby

주관식 문제

1. 데이터 사이언스를 간단하게 정의하시오.

- 데이터 사이언스는 통계 방법을 사용하여 대량의 데이터를 처리하는 여러 학문 분야가 관련된 과학으로서, 데이터에 대한 직관력을 획득할 수 있게 해주는 학문 분야이다.

2. 데이터 사이언스는 어떤 분야들의 융합으로 이루어지는지 말하시오.

- 데이터 사이언스는 인공지능, 컴퓨터 사이언스, 통계적 수학, 패턴인식, 정보공학, 머신러닝, 데이터베이스, 빅데이터, 데이터 마이닝, 그리고 비즈니스 등 다양한 분야들이 융합된 학문 영역이라 볼 수 있다.

3. 사용자가 빅데이터를 활용하는 단계를 간단히 나열하시오.

- 기업의 비즈니스 요구사항을 확인한다.

- 필요한 빅데이터를 검색하여 수집한다.

- 수집한 데이터를 적절한 형태로 가공한다.

- 처리된 데이터를 분석하고 시각화하여 사용한다.

4. 데이터 사이언스가 비즈니스에 미치는 좋은 효과를 말하시오.

- 의사 결정의 향상

- 경영상의 문제 개선

- 미래에 대한 예측

- 경영 목표 개선

- 새로운 기회 창출

- 위험 평가 향상

- 데이터 기반 습관

5. 데이터 마이닝이 많이 활용되는 기업 관련 업무를 3가지 정도 적으시오.

- 기업의 생산 과정에서 불량률을 줄이는 품질관리 분야

- 패턴인식 기법을 적용한 의료 진단 분야

- 고객의 신용을 평가하는 금융 관리 분야