

8.2

빅데이터의 개념 및 특성

- ➔ 빅데이터 분석은 기존 데이터 분석에 비해 (1) 규모가 매우 큰 데이터를 대상으로 하며, (2) 로그데이터(log data)나 구매기록 등 정형데이터뿐 아니라 소셜 미디어, 위치, 센서 등 반정형·비정형데이터까지 분석대상에 포함함
- ➔ 빅데이터는 (3) 다양한 데이터의 관계를 동시에 가능한 빨리 처리할 수 있는 새로운 컴퓨팅 기술을 적용하며, (4) 다양하고 신뢰할 수 있는 분석결과를 제시하여 가치를 창출하는 데이터 처리방식
- ➔ 빅데이터는 최소한 '5V'로 대표되는 규모(Volume), 다양성(Variety), 속도(Velocity), 정확성(Veracity), 가치(Value)의 5가지 구성요소를 갖추어야 함

8.2

빅데이터의 개념 및 특성

[표 8-3] 빅데이터의 5가지 구성요소(5V)

구분	주요내용
규모(Volume)	- 기술적인 발전과 IT일상화가 진행되면서 해마다 디지털 정보량이 기하급수적으로 폭증
다양성(Variety)	- 로그 기록, 소셜, 위치, 센서 데이터 등 데이터 종류의 증가 (반정형, 비정형데이터의 증가)
속도(Velocity)	- 소셜 데이터, IoT 데이터(센서, 모니터링), 스트리밍 데이터 등 실시간성 데이터 증가 - 실시간성으로 데이터 생성, 이동(유통) 속도의 증가
정확성(Veracity)	- 빅데이터의 특성상 방대한 데이터들을 기반으로 분석 수행 - 데이터 분석에서 고품질 데이터를 활용하는 것이 분석의 정확도(예측정확도)에 영향을 줌
가치(Value)	- 빅데이터가 추구하는 것은 가치 창출 - 빅데이터 분석을 통해 도출된 최종 결과물은 기업이 당면하고 있는 문제를 해결하는데 통찰력 있는 정보 제공

8.3

빅데이터 분석기법

8.3.1 기계학습(machine learning)

- ➔ 기계학습(machine learning, 머신러닝)이란 방대한 데이터를 분석해 미래를 예측하는 기술로 일반적으로 생성(발생)된 데이터를 정보와 지식(규칙)으로 변환하는 컴퓨터 알고리즘(algorithm)을 의미함
- ➔ 기계학습은 수집된 다양한 데이터 분석을 할 수 있는 기준(알고리즘)을 가지고 학습을 통해 해결책 제시(의사결정지원)를 자동화하는 것을 의미함
- ➔ 기계학습은 다양한 확률이론과 수학적 최적화기법, 통계기법, 컴퓨터 구조를 활용하여 이상적인 학습모델을 구축하는 기술과 연구자의 경험적 지식 습득까지 포함하는 융합기술

8.3

빅데이터 분석기법

8.3.2 빅데이터 분석기법: 데이터마이닝(Data Mining)

- ➔ 데이터마이닝 기법(Data Mining Techniques)은 축적된 대용량 데이터를 통계기법 및 인공지능기법을 이용하여 분석하고 이에 대한 평가를 거쳐 일반화시킴으로써 새로운 자료에 대한 예측 및 추측을 할 수 있는 의사결정을 지원함
- ➔ **데이터마이닝**은 대규모로 저장된 데이터 안에서 주로 인공지능기법을 활용하여 전통적인 통계학 이론으로는 설명이 힘든 패턴(pattern)과 규칙(rule)을 발견함
- ➔ 데이터마이닝은 분류(classification), 추정(estimation), 예측(prediction), 유사집단화(affinity grouping), 군집화(clustering)의 **5가지 업무영역**을 수행함

8.3.2 빅데이터 분석기법: 데이터마이닝(Data Mining)

업무영역	데이터마이닝 기법	연구 사례
분류(classification)	의사결정나무분석, 사례기반추론	부도예측(부도, 건전기업), 신용등급예측(우량, 불량고객), 보험사기적발(정상청구, 허위청구)
추정(estimation)	회귀분석, 신경망	배당금 산출, 고객평생가치(CLTV) 산출
예측(prediction)	장바구니분석, 사례기반추론, 의사결정나무, 신경망	소비자 구매행동 예측, 부도확률 예측, 환율변동성 예측 고객 이탈률 예측, 환자질병예측
유사집단화 (affinity grouping)	장바구니분석, 연관성분석	장바구니분석, 매장진열방법 교차판매, 끼워팔기 전략 수립
군집화(clustering)	클러스터링	시장 세분화, 고객세분화

- CRISP-DM(Cross-Industry Standard Process for Data Mining)은 데이터 분석 초보자도 사용할 수 있는 포괄적인 데이터마이닝 표준방법론이며, 공개표준프로세스임
- CRISP-DM은 (1) 비즈니스 이해 및 데이터 이해, (2) 데이터 준비, (3) 모델링 단계, (4) 모형 평가, (5) 모형 구축(개발)의 5단계로 구성됨
- 첫째, **비즈니스 이해 및 데이터 이해** 단계는 해당 비즈니스의 이해와 현업이 보유하고 관리하고 있는 데이터를 이해하는 단계
- 둘째, **데이터 준비** 단계는 자료를 컴퓨터 서버로부터 내려 받고 나서 분석 가능한 상태로 만들기 위해 데이터 정제(data cleaning) 작업을 하는 단계

- 셋째, **모델링** 단계는 자료 기술(data description) 및 탐색(exploration)을 포함하여 필요한 각종 모델링을 수행함,
여기에는 신경망, 의사결정나무 등의 지도학습과 군집화, 연관성분석 등의 비지도학습이 포함됨
- 넷째, **평가**는 앞 단계에서 생성된 모형이 잘 해석되는지, 독립적인 새로운 자료에 적용되는지 예측력 성과 등을 측정하는 단계
- 마지막, **모형 개발** 단계는 검토가 끝난 모형을 실제 현업에 적용하는 단계

8.3

빅데이터 분석기법

8.3.3 빅데이터 분석기법: 텍스트마이닝(Text Mining)

- ➔ 텍스트, 이미지, 음성데이터와 같이 정형화 되지 않은 비정형데이터 (unstructured data)를 다루는 기술이 빠르게 발전하고 있음
- ➔ 기업에서 생산되는 데이터의 80%이상은 비정형데이터로 이루어져 있으며, 그 중 텍스트데이터는 가장 대표적인 비정형데이터
- ➔ 온라인 쇼핑몰에서 사람들은 물건을 구매할 때 다른 구매자가 남긴 제품리뷰 텍스트(구매후기)로부터 제품에 대한 정보를 수집함
- ➔ 소셜네트워크서비스(SNS)에서 방대한 텍스트데이터(댓글데이터, 트윗글)가 생산되고, 빠르게 확산되고 있으며, 기업들은 이를 분석하여 마케팅 전략에 활용함

8.3

빅데이터 분석기법

8.3.3 빅데이터 분석기법: 텍스트마이닝(Text Mining)

➔ 텍스트마이닝은 자연어(natural language) 형태로 구성된 비정형 또는 반정형 데이터에서 패턴 또는 관계를 추출하여 의미 있는 정보를 찾아내는 기법으로 **자연어처리(natural language processing, NLP)가 핵심기술**

➔ 자연어처리(NLP)는 인공지능의 주요 분야 중 하나로 컴퓨터를 이용해 사람의 자연어를 분석하고 처리하는 기술

➔ 텍스트마이닝 분석을 실시하기 위해서는 불필요한 정보를 제거하고, 비정형 데이터를 정형데이터로 구조화하는 작업이 필요함,
데이터 불필요한 기호나 단어를 제거하는 토큰화(tokenization), 의미 없는 단어나 연구 주제와 일치하지 않는 단어를 제거해주는 불용어(stopword) 제외과정(관사, 전치사, 조사, 접속사 제외) 등

8.3

빅데이터 분석기법

8.3.3 빅데이터 분석기법: 텍스트마이닝(Text Mining)

- ➔ 텍스트마이닝 관련 분야로 감성분석(sentiment analysis), 오피니언 마이닝(opinion mining)
- ➔ 감성분석은 텍스트마이닝 분석의 한 분야로 특정 문서의 긍정, 부정에 대한 감정을 추측하고 분류하는 방법
- ➔ 사용자가 생성한 온라인 텍스트 속에 담긴 감성, 정서, 주관, 또는 감정을 식별하기 위해 사용됨
- ➔ 감성분석은 포럼(forum), 블로그(blog), SNS에서 발생하는 텍스트에 적용이 가능하며, 주로 소셜미디어 사용자들의 의견을 긍정(positive), 부정(negative), 중립(neutral)의 선호도로 판별하고, 감정을 추측 및 분류함

8.3

빅데이터 분석기법

8.3.4 빅데이터 분석기법: 소셜네트워크분석(Social Network Analysis: SNA)

- ➔ 소셜네트워크분석(Social Network Analysis: SNA)은 소셜시스템(social system)이 관계(relationship)와 이러한 관계에 의해 형성되는 패턴(pattern)에 의해 창조된다는 전제로 함
- ➔ SNA는 그래프 이론(Graph Theory)을 이용하여 사람, 그룹, 데이터 등 객체 간의 관계 및 관계 특성을 분석하고 시각화하는 측정기법
- ➔ 그래프는 점(node)과 선(link)으로 표현하며 점은 행위자를, 선은 행위자들 간의 관계를 표현함, 그래프 이론의 점과 선의 조합을 통해 사회적 관계(거래, 의사소통, 상호침투 등)를 표현함
- ➔ SNA는 사회 구성원 간의 관계에 분석의 초점을 맞추어 이들 관계의 패턴에서 의미 있는 시사점을 도출함

8.3

빅데이터 분석기법

8.3.4 빅데이터 분석기법: 소셜네트워크분석(Social Network Analysis: SNA)

- ➔ 최근에 수행되는 SNA는 트윗(tweet)의 확산분석 및 영향력자 분석(influencer analysis)에 집중됨
- ➔ 트윗의 확산분석은 특정 트윗의 확산유형을 네트워크 그래프로 분석(network graph analysis)하는 기술
- ➔ 영향력자(인플루언서, influencer) 분석은 트위터를 사용하는 사람들 중 특정 주제에 대해서 가장 영향력이 큰 사람을 순위화하는 기술
- ➔ 영향력자 분석은 트윗의 확산분석과 밀접하게 연관되며, 개별 트윗들의 확산분포를 분석하여 해당 트윗을 작성한 트위터리안(트위터 사용자)이 영향력자인지 여부를 결정함

8.4

빅데이터 분석단계: 과제기획과 과제분석단계

(1) 빅데이터 과제기획단계

➔ 빅데이터 과제기획단계는 (1) 문제 발굴 및 정의, (2) 문제해결을 위한 개념적 대안설계, (3) 데이터 가용성 평가, (4) 문제해결을 위한 논리적 모형설계, (5) 과제 추진방안 수립 및 타당성 평가, (6) 과제확정 및 분석계획수립으로 구성됨

➔ 첫째, 문제 발굴 및 정의는 빅데이터 분석을 통해서 향상된 고객서비스 제공 및 합리적·과학적 의사결정지원을 위한 기회 식별과 경영목표 달성을 위해 해결해야 할 문제를 식별하는 단계

➔ 둘째, 문제해결을 위한 개념적 대안 설계는 도출된 여러 가설 중 분석을 위해 필요한 가설을 추려내는 과정,

가설검정은 본격적인 데이터 분석을 위한 사전적인 대안설계 작업이며 식별된 가설검정을 위해 샘플데이터(sample data)를 수집함

8.4

빅데이터 분석단계: 과제기획과 과제분석단계

(1) 빅데이터 과제기획단계

- ➔ 셋째, **데이터 가용성 평가**는 관련된 데이터 존재 여부와 데이터 확보여부를 검토하는 단계
- ➔ 넷째, **문제해결을 위한 논리적 모형설계**는 논리적 모형과 필요한 변수를 선정하고 문제해결 대안을 수립함
- ➔ 다섯째, **과제 추진방안 수립 및 타당성 평가** 단계는 본격적으로 빅데이터 분석과제(project) 추진방안을 다양하게 검토함
- ➔ 마지막, **과제확정 및 분석계획수립** 단계는 여러 대안 중에서 평가과정을 거쳐 가장 우월한 대안을 선택하여 이를 과제(project)화하고, 계획단계의 입력물로 설정함,
빅데이터 분석기획단계에서 최종 선정된 프로젝트를 어떻게 수행하여 소기의 목적을 달성할 것인가에 대한 계획을 수립함

8.4

빅데이터 분석단계: 과제기획과 과제분석단계

(1) 빅데이터 과제기획단계

빅데이터 분석단계	세부단계	역할 및 활동
빅데이터 과제기획단계	문제 발굴 및 정의	- 요구사항 분석 - 가설 설정
	문제해결을 위한 개념적 대안 설계	- 가설 검정 - 샘플데이터 수집
	데이터 가용성 평가	- 관련 데이터 존재 여부 파악 - 데이터 확보 여부 검토
	문제해결을 위한 논리적 모형 설계	- 논리적 모형과 변수 선정 - 문제해결대안수립
	과제 추진방안수립 및 타당성 평가	- 경제적 타당성 분석, - 기술적 타당성 분석
	과제확정 및 분석계획수립	- 프로젝트 목표 정의 - 프로젝트 수행 예산 수립 - 프로젝트 관리 계획 수립

(2) 빅데이터 과제분석단계

→ 과제가 기획되고 추진계획이 수립되면 그 계획에 의거 과제분석을 수행하게 됨, **과제분석단계**는 (1) 데이터 수집, (2) 데이터 전처리와 정제, (3) 데이터 분석과 정리 및 처리결과의 수용, (4) 해석과 결과제시로 구성

→ 첫째, **데이터 수집**은 선정된 변수에 의해 구성된 분석모형이나 과제를 해결하기 위해 관련 데이터를 수집하고 이를 분석하는 단계

→ 둘째, **데이터 전처리(data preprocessing)**와 **정제(refinement)**는 다양한 소스(데이터 원천)로부터 획득한 데이터 중 분석하기에 부적합하거나 수정이 필요한 경우 데이터를 전처리, 정제하는 과정,

빅데이터 과제분석단계에서 데이터 전처리와 정제 과정은 프로젝트의 90% 이상을 차지함

8.4

빅데이터 분석단계: 과제기획과 과제분석단계

(2) 빅데이터 과제분석단계

→ 셋째, **데이터 분석과 정리 및 처리결과의 수용**은 분석 대상과 관련된 변수의 데이터가 수집되면 이를 분석하는 과정,

데이터 분석이란 모아 놓은 데이터에서 변수들 간의 관련성을 파악하는 것

→ 넷째, 분석의 마지막 단계는 **분석결과의 의미를 제시**하는 단계,

데이터 분석을 통해 변수 간의 관련성이 파악되면 그 결과가 의미하는 바를 명료하게 해석하여 의사결정자에게 구체적인 조언을 함

8.4

빅데이터 분석단계: 과제기획과 과제분석단계

(2) 빅데이터 과제분석단계

빅데이터 분석단계	세부단계	역할 및 활동
빅데이터 과제분석단계	데이터 수집	관련 데이터 수집
	데이터 전처리와 정제	데이터 전처리 및 정제 실시
	데이터 분석과 정리 및 처리결과 수용	데이터 분석 실시, 처리결과 수용 여부 결정, 모형의 정확성 향상
	해석과 결과제시	의사결정자에게 조언, 시각화 도구 활용

- ➔ 빅데이터 환경에서 기업은 경쟁우위 확보를 위해 빅데이터 분석을 수행하고 있으며, 성공적인 빅데이터 분석을 위해 체계적인 분석과정과 표준방법론을 지속적으로 개발하고 있음
- ➔ 빅데이터 분석 프로젝트 개발방법론(Big Data Analysis Project Development Methodology, 빅데이터 분석방법론)은 정보전략기획(ISP), 비즈니스 인텔리전스(BI) 방법론, 데이터웨어하우스(data warehouse) 구축방법론, 소프트웨어 개발방법론 등을 참조하여 개발됨
- ➔ 빅데이터 분석방법론은 문제정의, 데이터준비, 모델설계, 모델구현, 결과평가, 서비스구현의 6단계로 구성됨

- ➔ (1) **문제정의단계(Problem Definition Stage)**는 발주기관의 사업적 핵심문제점을 발견하고, 이들 문제점을 해결한 경우의 사업적 가치를 파악함(환경분석, 현황분석, 목표정의 수행)
- ➔ (2) **데이터준비단계(Data Preparation Stage)**는 프로젝트 수행 기간 동안 분석작업을 수행할 시스템을 구축하고, 분석환경 마련함, 현업이 보유 및 관리하고 있는 데이터를 이해하고 레코드의 수, 변수의 종류, 자료 값의 질(quality), 데이터 관리체계 등을 파악함
- ➔ (3) **모델설계단계(Model Design Stage)**는 다양한 데이터 소스로부터 수집된 테이블의 속성들 간의 관계를 파악하여 업무흐름과 데이터에 대한 개념적 이해를 명확히 함, 분석목적에 해당하는 속성(반응변수, 종속변수)과 상관관계가 높은 속성(설명변수, 독립변수)을 선별함

- (4) **모델구현단계(Model Development Stage)**는 분석모델을 실행하기 위해 학습용 데이터 집합, 검증용 데이터 집합, 그리고 테스트 데이터 집합으로 구성, 분류, 추정, 예측, 유사집단화, 군집화를 위한 신경망, 의사결정나무, 연관성분석, 통계기법 등의 다양한 데이터마이닝 기법을 이용하여 분석모델(연구모형)을 구축함
- (5) **결과도출단계(Result Extraction Stage)**는 분석모델을 실행하여 도출된 최종결과물을 점검하고, 사업적 측면에서 결과의 가치를 재평가
- (6) **서비스구현단계(Service Development Stage)**는 분석모델을 파일럿 테스트(pilot test)를 통해 운영한 다음 안정적으로 확대하여 운영계 시스템에 구축함

세부 단계	역할 및 활동
1단계: 문제정의단계	<ul style="list-style-type: none"> - 문제 인식 및 요구사항 분석, 가설 설정 - 환경분석, 현황분석, 목표정의 - 프로젝트 계획 수립
2단계: 데이터준비단계	<ul style="list-style-type: none"> - 분석 환경 마련, 데이터 이해 - 데이터 정제 및 통합, 속성 도출과 데이터 형식 적용
3단계: 모델설계단계	<ul style="list-style-type: none"> - 독립변수와 종속변수 설정 - 분석 모델(모형화기법) 선택 - 분석 모델의 실행 가능성 점검
4단계: 모델구현단계	<ul style="list-style-type: none"> - 학습용, 검증용, 테스트 데이터 구성 - 분석 모델(연구모형) 구축, 모형 평가 및 프로세스 검토
5단계: 결과도출단계	<ul style="list-style-type: none"> - 최종 결과물 점검, 사업적 가치 판단 - 최종 결과물 발표 및 완료보고서 작성
6단계: 서비스구현단계	<ul style="list-style-type: none"> - 파일럿 서비스, 운영계 시스템 구축 - 프로젝트 최종 산출물 배포, 유지계획 수립(모니터링)

Q&A

빅데이터 생태계란 무엇인가?

- ➔ **빅데이터 생태계**는 데이터 생산자(기업 또는 개인), 유통자(플랫폼), 소비자, 솔루션 제공자간 참여와 협력을 위한 일종의 네트워크
- ➔ 빅데이터 생태계는 서비스 사용자, 서비스 공급자, 그리고 애플리케이션 공급자로 구성됨
- ➔ 빅데이터 생태계의 구성요소: 기술, 데이터베이스, 데이터 제공자, 분석인프라, 운영인프라, 서비스인프라, 서비스데이터, 비즈니스 인텔리전스, 시각화도구, 응용프로그램 등(상세화)
- ➔ 빅데이터는 데이터 생성, 수집, 분석, 소비에 이르는 일련의 가치사슬을 통해 문제를 해결하는데 이러한 가치사슬은 구성원 간 협력과 새로운 가치를 창출하기 위한 가치네트워크(value network)로 발전되고 궁극적으로 가치네트워크가 결합하면 빅데이터 생태계가 구축됨

8.6

분야별 빅데이터 활용사례

8.6.1 고객(CRM)빅데이터 분석을 통한 구매성향예측

- ➔ CRM 분야는 빅데이터 분석이 가장 많이 활용되는 분야로 고객빅데이터(CRM빅데이터)분석을 통해 차별화된 고객서비스 제공함
- ➔ 고객의 서비스 이용성향 및 취향 등을 분석하여 고객 세분화(segmentation)와 각 고객 군에 맞는 맞춤형 전략을 추진하고, 고객의 불만사항을 실시간으로 파악하여 고객 충성도를 증진시킴
- ➔ 충성도 높은 고객뿐 아니라 이탈 가능성 높은 고객과 잠재 고객 파악에도 고객빅데이터 분석이 활용될 수 있음
- ➔ 미국의 마케팅 기업 카탈리나 마케팅 코퍼레이션(Catalina Marketing Corporation, 카탈리나 사)은 CRM빅데이터 분석을 통해 고객 구매성향을 예측하여 고객 니즈에 맞는 제품과 서비스 제공

8.6

분야별 빅데이터 활용사례

8.6.1 고객(CRM)빅데이터 분석을 통한 구매성향예측

- ➔ 다양한 데이터마이닝 기법(인공지능기법)을 이용하여 실시간으로 고객의 구매성향을 예측함, 고객에게 맞춤형 쿠폰 발행
- ➔ 카탈리나 사는 CRM 빅데이터를 활용하여 구매예측모형의 예측성과 향상과 더불어 맞춤형 고객 서비스를 제공하고 있음
- ➔ 최근 CRM 빅데이터 분석은 향상된 고객 맞춤형 서비스를 위해 추천시스템(recommendation systems) 개발에도 적용됨
- ➔ 아마존(Amazon)은 모든 고객들의 구매 내역을 데이터베이스에 기록하고 이를 분석하여 소비자의 소비 패턴과 관심사를 파악함, 아마존은 빅데이터를 활용한 항목(item, 상품) 기반 협업 필터링(item-to-item collaborative filtering) 추천시스템 개발함

8.6 분야별 빅데이터 활용사례

8.6.2 금융분야의 빅데이터 분석 및 활용

- ➔ 금융분야의 가장 대표적인 업종인 은행업종은 대출의사결정과 관련하여 해당 기업에 대한 도산예측(bankruptcy prediction)과 개인의 신용등급예측(credit rating prediction)에 금융빅데이터 분석을 적극 활용함
- ➔ 보험업종은 해약 가능성이 높은 고객, 고객가치가 높은 고객, 금융리스크가 높은 고객 등을 추정하기 위해 빅데이터 분석 활용, 보험 계약 및 보험금 심사 등의 업무에도 빅데이터 분석 활용
- ➔ 신용카드업종은 신용카드의 도용 패턴을 탐색하기 위해 빅데이터 분석을 이 용함, 신용카드 사용 이력(소비 성향 패턴)을 분석하여 고객의 니즈를 파악하고, 선호도가 가장 높은 품목에 할인 혜택을 제공함

8.6 분야별 빅데이터 활용사례

8.6.3 의료분야의 빅데이터 분석 및 활용

- ➔ 의료분야는 각종 임상 및 검진의료기기(device) 데이터의 통합, 병원 간 검사 기록과 연구데이터 공유를 통해 질병의 징후를 조기에 발견하고 질병을 예측하는데 빅데이터 도입과 활용이 확대되는 추세
- ➔ 미국 세튼 헬스케어 패밀리(Seton Health Care Family) 재단은 연간 200만 명에 달하는 환자들의 복잡한 진료 정보를 IBM의 왓슨(Watson) 시스템을 도입하여 환자가 미래에 겪을 수 있는 질환을 예측하여 이를 예방할 수 있는 통합형 빅데이터 의료서비스를 제공함
- ➔ IBM의 인공지능 의사 왓슨은 매일 쏟아지는 300개 이상의 의학저널, 200개 이상의 의학교과서, 1천500만 페이지에 달하는 의료정보, 치료 가이드라인과 전세계 암환자 2만 명의 치료사례를 분석해 각 환자에게 최적의 치료법을 제안함

8.6 분야별 빅데이터 활용사례

8.6.3 의료분야의 빅데이터 분석 및 활용



[그림 8-1] 인공지능 의사 왓슨(AI Watson) (출처: 계명대 동산병원)

8.6

분야별 빅데이터 활용사례

8.6.3 의료분야의 빅데이터 분석 및 활용

- ➔ 미국 국립암연구소(National Cancer Institute)는 암 환자에 대한 빅데이터 분석을 실시하여 암 발생의 원인을 제시함
- ➔ 환자의 상태와 가족력 등을 분석하여 이를 고려한 식단, 운동, 치료가 가능한 맞춤형 의료서비스를 제공하고 있음
- ➔ 빅데이터가 의료분야에 응용되면서 의료 패러다임이 바뀌고 있음, 의료는 병에 걸리면 치료를 한다는 치료의 개념이 강했지만 빅데이터 분석을 활용하면 질병을 원천적으로 차단하는 **예방 의료**가 가능해짐

8.6 분야별 빅데이터 활용사례

8.6.4 빅데이터 분석을 통한 재난관리

- ➔ 현재 미국, 일본 등 재난관리 선진국은 재난관리 분야에 빅데이터를 적극적으로 활용함
- ➔ 환경 센서(sensor) 데이터를 이용한 지진 감지, SNS를 활용한 현황 파악 및 복구 지원, 위성데이터를 활용한 재난 지역 파악 등을 빅데이터 분석을 통해 수행하고 있음
- ➔ 미국 국립해양대기청(NOAA)은 빅데이터 분석을 통해 기상정보 및 지질정보를 추정하고 있음
- ➔ 기온과 상대습도의 조합으로 계산되는 열지수 강도와 지속시간을 기준으로 여름철 폭염특보와 고온건강경보시스템 등을 제공함,
NOAA은 빅데이터를 기반으로 신경망과 의사결정나무 등의 인공지능기법을 이용하여 기상예측모형을 구축함

8.6 분야별 빅데이터 활용사례

8.6.4 빅데이터 분석을 통한 재난관리

- ➔ 빅데이터를 이용하여 가장 예측하기 힘든 분야는 지진과 쓰나미, 과거의 재해 빅데이터와 현재 발생지진 정보를 결합·분석하여 지진과 쓰나미 등의 자연재해를 예측하기 위해 노력하고 있음
- ➔ 미국지질조사국(USGS)은 1900년 이래로 발생한 각종 지진을 유형별, 크기별로 조사할 뿐 아니라 그 피해정도까지 분석하여 지진이 발생한 경우 어떤 형태로 발전할지 재난 상황을 미리 시뮬레이션 함으로써 각종 재난 피해를 예측함
- ➔ 최근에는 다양한 기계학습 및 인공지능 기반의 예측성고가 뛰어난 자연재해 예측모형을 구축하여 미래 재난에 대처할 수 있는 기반을 마련함

8.6

분야별 빅데이터 활용사례

8.6.5 범죄예방 및 수사에서의 빅데이터 분석 및 활용

- 미국, 영국, 독일, 싱가포르 등에서 범죄예방 및 수사에 빅데이터를 적극 활용하고 있음
- 2013년 보스턴 마라톤 테러 용의자를 신속하게 검거할 수 있었던 배경에는 빅데이터 분석을 통해 용의자 추적 기술을 활용한 것
- 테러직후 주변 600여대의 CCTV 데이터, 시민들이 현장에서 촬영한 사진과 영상데이터, 트위터(twitter)와 페이스북(facebook)의 글, 그리고 현장 기지국 통화 로그데이터(log data) 등의 공공빅데이터를 분석하여 용의자를 식별함
- 미국 국세청(Internal Revenue Service)은 2011년, SAS사와 공동으로 빅데이터와 인공지능기법(신경망 등) 기반의 통합형 탈세 및 사기범죄방지시스템을 구축함

8.6 분야별 빅데이터 활용사례

8.6.5 범죄예방 및 수사에서의 빅데이터 분석 및 활용

➔ 로스앤젤레스 경찰국(LAPD)은 과거에 발생한 범죄수법, 시·공간 환경적 통계정보를 분석하여 실시간 범죄 지도를 구성하고, 범죄 발생을 미연에 방지하는 범죄예측서비스(프레드폴, PredPol) 운영 중

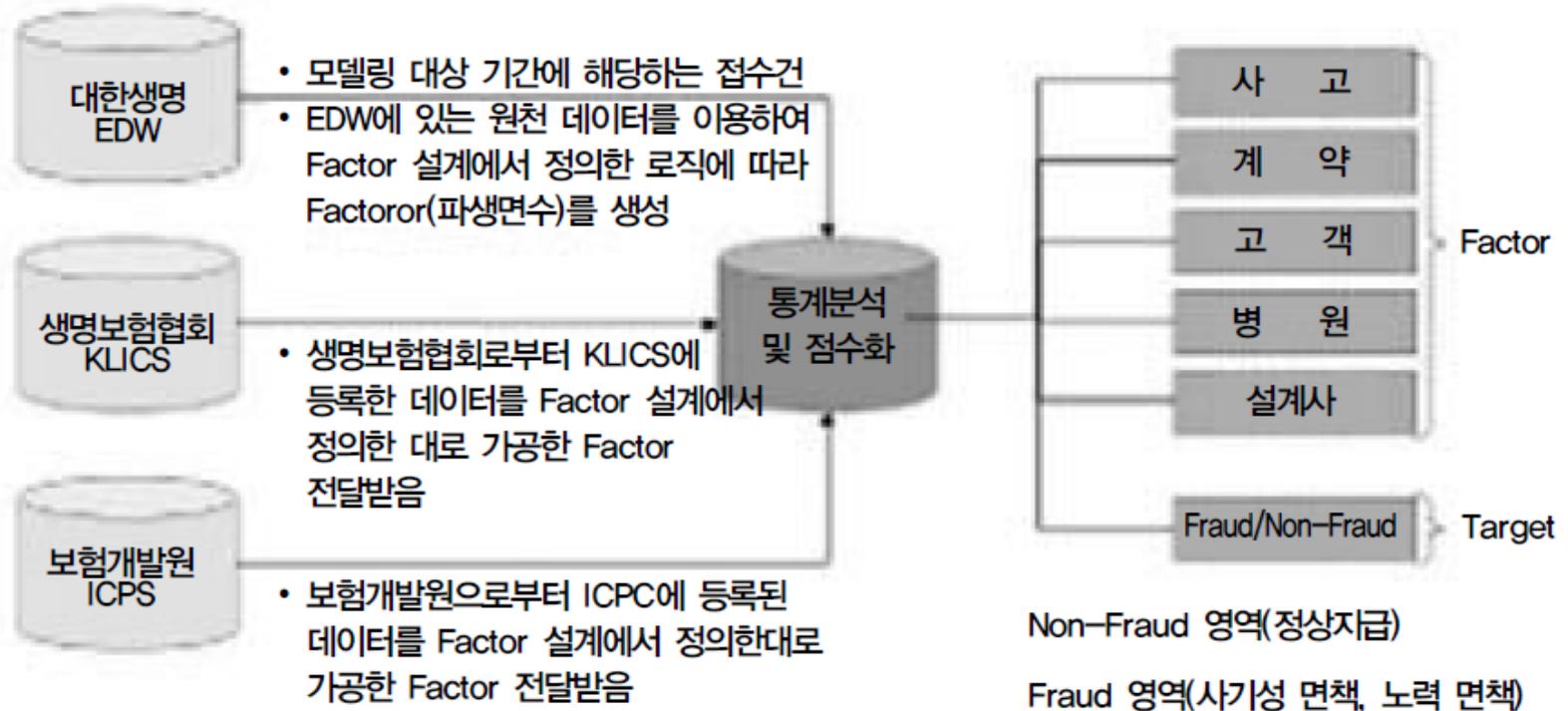
➔ 과거 범죄 통계 정보와 우범지역 CCTV 영상에서 범죄가 일어나는 상황을 인공지능이 학습하고, 이를 통해 범죄발생 위험이 높은 장소와 시간을 예측하여 경찰관들의 범죄예방을 위한 직관을 보완해 줌

➔ 뉴욕 경찰청(NYPD)은 2012년 마이크로소프트(MS)사와 공동으로 최첨단 범죄감시시스템인 Domain Awareness System(DAS)을 450억 원의 개발비용으로 구축함

➔ DAS를 활용하면 위치정보를 통해 범죄용의자 차량이 현재 어디에 있는지, 과거 어느 지역에 있었는지 추적 가능함

- ➔ 최근 기업들은 빅데이터를 활용한 부정행위 관리를 통해 사업의 무결성(integrity)을 확보하고 있음
무결성 확보는 부정 및 사기방지가 주요 업무이며, 빅데이터를 활용하여 사기 의심 거래 후보군을 정밀하게 예측하고 있음
- ➔ 대한생명(현 한화생명보험)은 2007년, **보험사기방지시스템(insurance fraud detection system, IFDS)**을 구축함
- ➔ IFDS는 보험빅데이터를 기반으로 예측분석을 통해 보험사고 허위사실이나 확대 청구 등 다양한 사기 행위들을 객관적이고, 현실적으로 적발하는 지능형 예측모델시스템
- ➔ IFDS는 보험사기로 적발된 사례들을 수집, 패턴화하여 보험사기 혐의자(사기 의심 거래 후보군)를 자동으로 추출함

➔ 대한생명은 자사 데이터만을 활용한다는 한계를 극복하고자 2010년, IFDS 을 고도화함, 자사 데이터뿐 아니라 보험개발원, 보험협회 등의 공공 빅데이터를 함께 활용할 수 있고, 딥러닝(심층학습)기법을 이용한 빅데이터 분석도 가능함



핵심용어 (1)

- 기계학습(machine learning, 머신러닝): 방대한 데이터를 분석해 미래를 예측하는 기술로 일반적으로 생성(발생)된 데이터를 정보와 지식(규칙)으로 변환하는 컴퓨터 알고리즘(algorithm)을 의미함
- 인공신경망(artificial neural network): 기계학습과 인지과학에서 생물학의 신경망에서 영감을 얻은 통계학적 학습알고리즘, 두뇌의 정보처리 과정을 모방함
- 데이터마이닝(data mining): 축적된 대용량 데이터를 통계기법 및 인공지능기법을 이용하여 분석하고 이에 대한 평가를 거쳐 일반화시킴으로써 새로운 자료에 대한 예측 및 추측을 할 수 있는 의사결정을 지원함
- 데이터마이닝 업무영역: 분류(classification), 추정(estimation), 예측(prediction), 유사집단화(affinity grouping), 군집화(clustering)

핵심용어 (2)

- CRISP-DM: 데이터마이닝 방법론으로 (1) 비즈니스 이해 및 데이터 이해, (2) 데이터 준비, (3) 모델링 단계, (4) 모형 평가, (5) 모형 구축(개발)로 구성됨
- 텍스트마이닝(text mining): 자연어 형태로 구성된 비정형 또는 반정형데이터에서 패턴 또는 관계를 추출하여 의미 있는 정보를 찾아내는 기법
- 자연어처리(natural language processing, NLP): 인공지능의 주요 분야 중 하나로 컴퓨터를 이용해 사람의 자연어를 분석하고 처리하는 기술로 자연어 분석, 자연어 이해, 자연어 생성의 기술이 사용됨
- 감성분석(Sentiment analysis): 소비자의 감성과 관련된 텍스트 정보를 자동으로 추출하는 텍스트마이닝(text mining) 기술의 한 영역, 주로 온라인 쇼핑몰에서 사용자의 상품평에 대한 분석이 대표적 사례임

핵심용어 (3)

- 소셜미디어(social media): SNS에 가입한 이용자들이 서로 정보와 의견을 공유하면서 대인관계망을 넓힐 수 있는 플랫폼
- 오피니언 마이닝(Opinion Mining): 소셜미디어 등 주로 비정형텍스트의 긍정(positive), 부정(negative), 중립(neutral)의 선호도를 판별하는 기술임, 특정서비스 및 상품에 대한 시장규모 예측, 소비자의 반응, 입소문 분석(viral analysis) 등에 활용됨
- 소셜네트워크분석(Social Network Analytics, SNA): 수학의 그래프 이론에 기반함, 소셜네트워크 연결구조 및 연결강도를 바탕으로 사용자의 명성 및 영향력을 측정하여 SNS상에서 입소문의 중심이나 허브(hub)역할을 하는 사용자를 찾는데 활용됨
- 영향력자(인플루언서, influencer)분석: 트위터를 사용하는 사람들 중 특정 주제에 대해서 가장 영향력이 큰 사람을 순위화(ranking)하는 기술

핵심용어 (4)

- ➔ 빅데이터 과제 기획 단계: (1) 문제 발굴 및 정의, (2) 개념적 대안설계, (3) 데이터 가용성 평가, (4) 논리적 모형설계, (5) 추진방안 수립 및 타당성 평가, (6) 과제확정 및 분석계획수립으로 구성
- ➔ 빅데이터 과제 분석 단계: (1) 데이터 수집, (2) 데이터 전처리와 정제, (3) 데이터 분석과 정리 및 처리결과의 수용, (4) 해석과 결과제시로 구성
- ➔ 빅데이터 분석 방법론: 문제정의, 데이터준비, 모델설계, 모델구현, 결과평가, 서비스 구현의 6단계로 구성됨
- ➔ 빅데이터 생태계: 데이터 생산자(기업 또는 개인), 유통자(플랫폼), 소비자, 솔루션 제공자간 참여와 협력을 위한 일종의 네트워크를 말함
- ➔ 보험사기방지시스템(insurance fraud detection system, IFDS): 보험빅데이터를 기반으로 예측분석을 통해 보험사고 허위사실이나 확대 청구 등 사기 행위들을 객관적이고, 현실적으로 적발하는 지능형 예측모델시스템

핵심용어 (5)

- 버즈마케팅(buzz marketing): 구전 마케팅(viral marketing)의 일종으로 상품을 이용해 본 소비자가 자발적으로 그 상품에 대해 주위 사람들에게 긍정적인 입소문을 퍼트리도록 유도하는 것
- 디지털 데이터 마케팅: 소비자가 온라인상에서 제품이나 서비스를 구매하기 위해 단순히 검색을 하는 것부터 구매 의사결정을 내리기까지 이루어지는 모든 흔적을 단계별로 분석하여 소비자의 구매 촉진에 영향을 주는 요인을 찾아 마케팅 인사이트(insight)를 도출하고 유의미한 전략을 창출하는 것
- 프레드폴(PredPol): 과거에 발생한 범죄수법, 시·공간 환경적 통계정보를 분석하여 실시간 범죄 지도를 구성하고, 범죄 발생을 미연에 방지하는 범죄예측서비스

연습문제 (1)

1. 다음 설명에서 ㉠과 ㉡가 무엇을 말하는지 답하십시오.

시장조사기관인 포레스터 리서치(Forrester Research, Inc.)의 “빅데이터 기술 수명주기 예측보고서”에서는 빅데이터 기술은 향후 10년간 안정적 성장세를 유지할 것이며, 비즈니스 가치가 높은 핵심기술이라고 언급하였다. 위 보고서에서는 상위 5개 빅데이터 기술로 (㉠), 빅데이터 처리(BigData preparation, NoSQL databases), (㉡), 인공지능(artificial intelligence), 기계학습(machine learning)을 제시하였다.

㉠ - () ㉡ - ()

2. 다음 설명에서 ㉠과 ㉡가 무엇을 말하는지 답하십시오.

광의의 빅데이터 개념에서는 빅데이터의 5가지 구성요소(5V)로 규모(Volume), 다양성(Variety), 속도(Velocity), (㉠), (㉡)를 제시하였다.

㉠ - () ㉡ - ()

연습문제 (2)

3. 다음 설명에서 ㉠가 무엇인지 답하시오.

(㉠)이란 방대한 데이터를 분석해 미래를 예측하는 기술로 일반적으로 생성(발생)된 데이터를 정보와 지식(규칙)으로 변환하는 컴퓨터 알고리즘(algorithm)을 의미한다.

㉠ - ()

4. 다음 설명에서 ㉠와 ㉡가 무엇을 말하는지 답하시오.

데이터마이닝은 (㉠), 추정(estimation), (㉡), 유사집단화(affinity grouping), 군집화(clustering)의 5가지 업무영역으로 구분할 수 있다.

㉠ - () ㉡ - ()

5. 다음 설명에서 ㉠와 ㉡가 무엇을 말하는지 답하시오.

CRISP-DM은 데이터마이닝 작업의 표준화 실행 단계로 비즈니스 이해 및 데이터 이해, 데이터 준비, (㉠), (㉡), 모형 구축(개발)의 5단계를 제시하였다.

㉠ - () ㉡ - ()

연습문제 (3)

6. 다음 설명에서 ㉠과 ㉡가 무엇을 말하는지 답하십시오.

자연어처리(NLP)는 인공지능의 주요 분야 중 하나로 컴퓨터를 이용해 사람의 자연어를 분석하고 처리하는 기술로 자연어 분석, (㉠), (㉡)의 기술이 사용된다. 최근에는 심층기계학습(deep learning, 딥러닝) 기술이 기계번역 및 자연어 생성에 적용되고 있다.

㉠ - () ㉡ - ()

7. 다음 설명에서 ㉠과 ㉡가 무엇을 말하는지 답하십시오.

빅데이터 분석단계는 빅데이터 과제기획단계와 빅데이터 과제분석단계로 구성된다. 먼저, 빅데이터 과제기획단계는 (1) 문제 발굴 및 정의, (2) (㉠), (3) (㉡), (4) 문제해결을 위한 논리적 모형설계, (5) 과제 추진방안 수립 및 타당성 평가, (6) 과제확정 및 분석계획수립으로 구성된다.

㉠ - () ㉡ - ()

연습문제 (4)

8. 다음 설명에서 ㉠가 무엇인지 답하시오.

과제가 기획되고 추진계획이 수립되면 그 계획에 의거 과제분석을 수행하게 된다. 빅데이터 과제분석단계는 (1) 데이터 수집, (2) (㉠), (3) 데이터 분석과 정리 및 처리결과의 수용, (4) 해석과 결과제시로 구성된다.

㉠ - ()

9. 다음 설명에서 공통적으로 ㉠와 ㉡가 무엇을 말하는지 답하시오.

빅데이터 분석방법론은 문제정의, 데이터준비, (㉠), (㉡), 결과평가, 서비스구현의 6단계로 구성된다. 특히 (㉠)와 (㉡)단계는 긴밀한 협조가 필요하므로 반복 수행되기도 한다.

㉠ - () ㉡ - ()

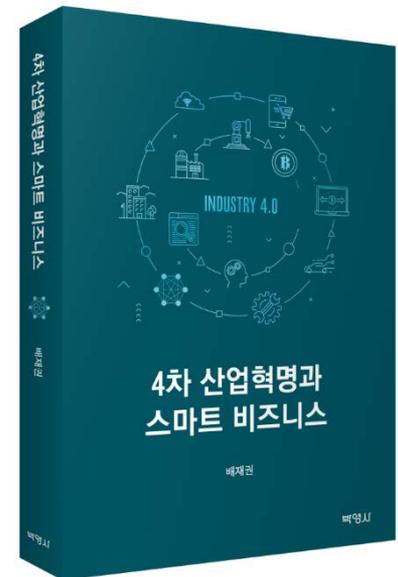
연습문제 (5)

10. 다음 설명에서 ㉠가 무엇인지 답하시오.

(㉠)은 보험빅데이터를 기반으로 예측분석을 통해 보험사고 허위사실이나 확대 청구 등 다양한 사기 행위들을 객관적이고, 현실적으로 적발하는 지능형 예측모델시스템이다.

㉠ - ()

[Chapter 09] 인공지능과 4차 산업혁명



- ➔ 4차 산업혁명이 3차 산업혁명과 구별되는 점은 **로봇과 인공지능(Artificial Intelligence, AI)**을 활용하여 자동적·지능적으로 제어할 수 있는 가상의 물리시스템(CPS) 구축이 가능하다는 것
- ➔ 4차 산업혁명에서 가장 중요한 화두는 바로 인공지능 기술의 발전과 적용(융합)
- ➔ 사물인터넷, 클라우드, 빅데이터, 그리고 인공지능이 융합되어 가상세계에서 예측과 맞춤을 통한 현실 세계를 최적화하는 O2O 플랫폼 구축에 인공지능은 중추적인 역할을 담당함
- ➔ **O2O 융합(convergence)**은 현실 세계에 물리적으로 실재하는 것과 사이버 공간의 데이터 및 소프트웨어를 실시간으로 통합,
O2O 융합의 핵심 기능은 예측과 맞춤이며, 이를 위해서 반드시 인공지능이 필요함

9.1

인공지능 기술발전의 역사

- ➔ 인공지능(Artificial Intelligence, AI)은 미국의 컴퓨터과학자인 존 매카시(John McCarthy)가 1956년 다트머스 컨퍼런스(Dartmouth Conference)에서 최초로 사용한 표현임
- ➔ 인공지능은 인간과 유사하게 사고하는 컴퓨터 지능을 일컫는 포괄적 개념으로 인간의 **지각**, **추론**, **학습능력** 등을 컴퓨터 기술을 이용하여 구현한 컴퓨터 알고리즘
- ➔ 존 매카시를 필두로 인공지능의 연구가 시작되었으며, 인공지능의 기술발전은 **계산주의 시대**, **연결주의 시대**, **딥러닝 시대**로 구분됨

- ➔ 초창기 시대는 **계산주의(computationalism)** 시대, 계산주의는 인간이 보유한 지식을 컴퓨터로 표현하고 이를 활용해 현상을 분석하거나 문제를 해결하는 지식기반시스템을 말함
- ➔ 1950년대 존 매카시를 비롯하여 마빈 민스키(Marvin Minsky), 나다니엘 로체스터(Nathaniel Rochester), 클로드 새넨(Claude Shannon) 등 당시 최고의 정보과학자들이 다트머스대학교(Dartmouth College)에 모여 계산주의 인공지능을 연구함
- ➔ 컴퓨팅 성능 제약으로 계산기능(연산기능)과 논리체계의 한계, 데이터 부족 등의 근본적인 문제에 직면하여 결국 계산주의 연구는 기대에 부응하지 못함

- ➔ 계산주의로 인공지능 발전에 제약이 생기면서 1980년대에 **연결주의 (connectionism)**가 새롭게 대두됨
- ➔ 연결주의는 지식을 직접 제공하기보다 지식과 정보가 포함된 데이터를 제공하고 컴퓨터가 스스로 필요한 정보를 학습함
- ➔ 연결주의는 인간의 두뇌를 모사하는 **인공신경망(Artificial Neural Network)**을 기반으로 한 모델,
연결주의 시대의 인공지능은 인간과 유사한 방식으로 데이터를 학습하여 스스로 지능을 고도화함
- ➔ 연결주의는 막대한 컴퓨팅 성능과 방대한 학습데이터가 필수적이거나 당시에 이들이 부족하여 비즈니스 활용 측면에서 한계가 있음,
연결주의 시대도 학습에 필요한 빅데이터와 컴퓨팅 파워의 부족이라는 한계를 극복하지 못함

- ➔ 2010년 이후 GPU(Graphic Processing Unit, 그래픽 프로세서)의 등장과 분산처리기술의 발전으로 계산주의와 연결주의 시대의 문제점인 방대한 양의 계산문제를 대부분 해결하게 됨
- ➔ 최근의 인공지능은 **딥러닝(deep learning, 심층학습)의 시대**, 딥러닝 역시 연결주의 시대와 동일하게 신경망을 학습의 주요 방식으로 사용함
- ➔ 신경망의 기본 구조인 입력층(input layer)과 출력층(output layer) 사이에 다수의 숨겨진 은닉층(hidden layer)으로 구성된 신경망을 딥 뉴럴 네트워크 (Deep Neural Networks, 심층신경망)라고 부르면서 딥러닝 용어가 탄생함

9.1

인공지능 기술발전의 역사 – 딥러닝 시대

- ➔ 최근에는 4차 산업혁명 기술력의 급성장, 혁신적 알고리즘의 등장, 그리고 분산컴퓨팅, 특히 GPU의 발전으로 딥러닝 기술이 진보하고 있음
- ➔ 현재 딥러닝은 음성인식, 이미지인식, 자동번역, 그리고 무인주행(자동차, 드론) 등에 큰 성과를 나타내고 있으며 의료, 법률, 세무, 교육, 예술 등 다양한 범위에서 활용됨

[표 9-1] 인공지능 발전의 시대적 구분

시대적 구분	시대	특징
계산주의 시대	1950년대	<ul style="list-style-type: none"> - 인간이 보유한 지식을 컴퓨터로 표현하고 이를 활용해 현상을 분석하거나 문제를 해결하는 지식기반시스템 - 계산기능(연산기능)과 논리체계의 한계, 데이터 부족 등
연결주의 시대	1980년대	<ul style="list-style-type: none"> - 지식을 직접 제공하기보다 지식과 정보가 포함된 데이터를 제공하고 컴퓨터가 스스로 필요한 정보를 학습하는 것 - 컴퓨팅 성능과 방대한 학습데이터가 부족하여 비즈니스 활용 측면에서 한계점 대두
딥러닝 시대	2010년 이후	<ul style="list-style-type: none"> - 컴퓨팅 성능 향상, GPU 등장, 분산처리기술의 발전 - 사물인터넷과 클라우드 기술의 발전으로 빅데이터 생성·수집됨 - 딥 뉴럴 네트워크(Deep Neural Networks, 심층신경망)의 활용 - 음성인식, 이미지인식, 자동번역, 무인주행(자동차, 드론) 등을 딥러닝에 적용

9.2

인공지능의 정의와 유형

- ➔ 아마존의 알렉사(Alexa), 애플(Apple)의 시리(Siri), MS의 코타나(Cortana) 등 대화형 에이전트(개인비서)에서 자율주행차의 인지·판단시스템에 이르기까지 인공지능 기술이 빠르게 확산되고 있음
- ➔ 미국의 미래학자 레이 커즈와일(Ray Kurzweil)은 2045년을 **기술적 특이점 (technological singularity)**, 특이점 기준 아래에 그 기준을 적용할 수 없는 점)으로 예상하면서 인간의 두뇌보다 창의력과 지혜, 사회적 능력 등이 뛰어난 초지능이 등장할 것으로 주장함
- ➔ 인공지능이 인간 지능에 의존하지 않고 스스로 진화해가면, 미래에 인간이 미래를 예측할 수 없는 시점에 이를 수도 있다는 것

9.2

인공지능의 정의

[표 9-2] 인공지능의 정의

연구자 및 기관	정의	핵심 키워드
Ray Kurzweil(1990)	인간에 의해 수행될 때 필요한 지능에 관한 기능을 제공하는 기계를 만드는 작업	지능, 기계
Rich & Knight(1991)	컴퓨터가 특정 순간에 사람보다 더 효율적으로 작업이 가능하도록 만드는 연구	효율성
Luger & Stubblefield(1993)	지능적인 행동의 자동화에 관한 컴퓨터 과학의 한 부문	지능자동화 컴퓨터과학
가트너(Gartner)	인간을 대체, 인지능력을 제고, 인간의 의사소통 통합, 복잡한 콘텐츠의 이해, 결론을 도출하는 과정 등 인간이 수행하는 것을 모방하는 기술	인지능력, 모방
한국정보화진흥원(NIA)	인간의 학습능력, 추론능력, 지각능력, 이해능력 등을 실현한 기술	학습, 추론, 지각, 이해능력
창조경제연구회(2016)	학습을 통하여 예측과 맞춤의 가치를 제공하는 최적화 시스템	학습, 예측, 맞춤, 최적화

→ 미국의 철학자 존 설(John Searle)은 인공지능은 주어진 조건에서 작동 가능한 약한 인공지능(**Weak AI** 또는 Artificial Narrow Intelligence)과 자의식을 지니고 인간과 같은 사고가 가능한 강한 인공지능(**Strong AI** 또는 Artificial General Intelligence)으로 분류

→ **약한 인공지능**은 인공지능 기술을 유용한 소프트웨어 기술로 파악하고, 특정 문제를 해결하기 위한 인간의 지능적 행동을 수행하도록 공학적 응용을 모색한 접근방식,

약한 인공지능은 합리적으로 생각하고, 행동하는 시스템

→ **강한 인공지능**은 인간과 같은 사고체계로 문제를 분석하고 행동할 수 있도록 인공지능을 연구하는 접근방식,

강한 인공지능은 인간처럼 생각하고 행동하는 시스템

[표 9-3] 약한 인공지능과 강한 인공지능

구분	약한 인공지능	강한 인공지능
인간의 사고	<p>합리적으로 생각하는 시스템</p> <ul style="list-style-type: none"> - 정신적 능력을 갖춘 시스템 - 사고의 법칙 접근 방식 	<p>인간처럼 생각하는 시스템</p> <ul style="list-style-type: none"> - 사고 및 의사결정을 내리는 시스템 - 인지 모델링 접근방식
인간의 행동	<p>합리적으로 행동하는 시스템</p> <ul style="list-style-type: none"> - 지능적 행동을 하는 에이전트 시스템 	<p>인간처럼 행동하는 시스템</p> <ul style="list-style-type: none"> - 어떤 행동을 기계가 따라하는 시스템

→ 인공지능을 에이전트(agent)로 인식하고 입력과 출력의 관계에서 그 수준에 따라 4단계의 레벨로 구분하는 방법

→ 첫 번째 레벨(Level 1)은 단순한 제어프로그램,
Level 1은 상업적으로 인공지능이라고 지칭하는 것들로 지극히 단순한 제어프로그램을 탑재한 전자제품(에이컨, 세탁기, 청소기 등)을 인공지능 탑재라고 부름

→ 두 번째 레벨(Level 2)은 고전적인 인공지능(Classical AI),
Level 2는 행동 패턴이 지극히 다채로운 경우의 지능을 말함,
장기프로그램, 퍼즐프로그램, 청소로봇, 진단프로그램, 그리고 단순한 질문에 대답하는 인공지능이 이에 해당함

9.2

인공지능의 수준(레벨)에 따른 분류

- ➔ 세 번째 레벨(Level 3)은 기계학습(machine learning)을 적용한 인공지능, Level 3은 검색엔진에 내장된 데이터 또는 빅데이터를 학습하여 자동적으로 판단하는 인공지능, Level 3 인공지능의 사례로는 추천시스템
- ➔ 네 번째 레벨(Level 4)은 딥러닝(deep learning)을 적용한 인공지능, 딥러닝(심층학습)은 기계학습을 실현하는 기술, 딥러닝은 다층구조 신경망(심층신경망, Deep Neural Network) 기반의 기계 학습 분야로, 다량의 데이터로부터 높은 수준의 추상화 모델을 구축하는 기법
- ➔ 현재 인공지능은 약한 인공지능에서 강한 인공지능으로의 전환과 딥러닝을 적용한 인공지능(레벨 4)으로 가속화되고 있음

- ➔ 최근 인공지능 선도기업들은 자사의 인공지능 알고리즘을 공개하여 인공지능 시스템의 표준화를 주도하고 있음
- ➔ 2015년을 기점으로 인공지능 소스코드를 **오픈소스(open source)**화하여 인공지능 개발에 필요한 투자비용을 대폭 감소시키고 있음
- ➔ 구글(Google)은 인공지능 오픈소스 텐서플로우(TensorFlow) 등 자사 보유의 인공지능 기술을 일반인에게 무료로 공개하고 있음
- ➔ 인공지능 개발기업과 연구자들은 공개된 오픈소스를 활용함으로써 인공지능 개발비용과 시간을 절감하고 있음
- ➔ 인공지능 소스코드 개방기업들은 빅데이터 플랫폼과 양질의 빅데이터를 보유하고 있다는 공통된 특징을 지니고 있음, 글로벌 기업들이 보유한 인공지능 기술이 공개됨에 따라 인공지능 생태계의 발전은 가속화되고 있음

- ➔ 공격적인 인수합병(Mergers & Acquisitions, M&A)을 통해 기술력과 인재를 확보하여 인공지능 시대를 준비하고 있음
- ➔ 인공지능 후발주자 기업은 M&A를 통해 자체 기술개발과 비교하여 개발 속도를 증진하는 것은 물론 신기술 융합을 가능하게 함
- ➔ 알파벳, 애플, 아마존, 삼성전자 등 글로벌 기업들은 경쟁적으로 스타트업(창업기업)을 인수하고 있으며 인수대상 기업은 대부분 4차 산업혁명 분야에 집중되고 있음
- ➔ 인공지능은 첨단기술 영역으로 어느 산업보다 기술개발을 이끌어갈 인재가 중요하기 때문에 공격적으로 인공지능의 석학들을 영입함
- ➔ 공격적으로 M&A을 통한 인재를 영입하는 등 인공지능 전문가 영입 경쟁은 계속해서 과열되고 있음

- ➔ 2015년 8월, 경영컨설팅기업 보스턴컨설팅그룹(BCG)과 경제전문잡지 포브스(Forbes) 등은 인공지능 분야를 포함한 글로벌 혁신기업을 선정하여 발표한 바 있음
- ➔ 글로벌 혁신기업 50군에는 미국이 29개, 일본이 5개 기업이 선정되었고, 국내 기업은 삼성전자, 네이버, 아모레퍼시픽이 선정됨
- ➔ 4차 산업혁명 기술(인공지능, 3D프린팅, 사물인터넷, 클라우드) 관련 분야에서 인공지능 소프트웨어 분야는 전 세계 158개 기업이 혁신기업으로 선정되었으나 국내는 단지 1개 기업만이 선정됨
- ➔ 국내의 인공지능 기술력은 선도국가인 미국에 비해 약 2년의 기술격차가 나고 있으며 이를 줄이기 위해 인공지능 관련 투자와 활성화 정책이 시급한 실정임

- 인간과 인공지능의 역사적 대결에서 인공지능은 모두 승리함
- 2016년 구글 딥마인드(Google DeepMind) 인공지능 바둑프로그램 알파고(AlphaGo)와 한국기원 소속의 프로기사 이세돌 9단의 대결은 알파고의 승리로 끝남
- 1997년 IBM의 딥 블루(Deep Blue)가 인간 체스 챔피언(가리 카스파로프, Garry Kasparov)과의 체스 게임에서 승리
- 2011년 IBM의 인공지능 왓슨(Watson)이 제퍼디(Jeopardy) 퀴즈쇼에서 인간 퀴즈 챔피언과 경쟁하여 우승
- 인간은 인공지능의 **학습능력**, **추론능력**, 그리고 **지각능력**을 넘을 수 없기에 인간과 인공지능의 대결은 의미가 없음

9.4

인공지능의 적용 분야 – 인간과 인공지능의 대결



[그림 9-1] IBM Watson과 인간과의 퀴즈대결(출처: www.ibm.com)

- ➔ 인공지능은 근본적으로 미래기술들과 융합하여 온라인과 오프라인이 융합하는 O2O 세상을 선도하고 있음
- ➔ O2O 융합을 이끄는 클라우드, 빅데이터, 사물인터넷, 생체인터넷, 핀테크(Fintech), 디지털 헬스케어(Digital healthcare) 등이 인공지능과 융합하여 혁신적인 비즈니스 모델과 서비스가 창출되고 있음
- ➔ 제조업·서비스업 분야에서 인공지능이 적용되어 자동화·지능화가 촉진되고, 생산성과 품질이 향상되고 있음
- ➔ 제조업의 스마트화, 스마트팩토리(smart factory, 지능형 공장)는 인공지능 기반의 사이버물리시스템을 통해 효율성을 높이고 있음

- ➔ 구글(Google)은 중·장기 인공지능 프로젝트의 일환으로 **무인자동차** 개발 프로젝트를 수행함
- ➔ 자동차에 부착된 센서(sensor)로 데이터를 확보하고, 클라우드에서 빅데이터를 축적하여 인공지능이 제공한 예측과 맞춤으로 가치를 창출하고 현실을 최적화시키는 과정을 무인자동차로 구현함
- ➔ 구글의 무인자동차는 부착된 센서를 통해 주변의 환경을 감지하면서 도로 위의 데이터를 수집함, 수집한 데이터는 클라우드를 통해 데이터로 축적(저장)되고, 이를 인공지능(딥러닝)으로 분석함
- ➔ 인공지능은 분석한 데이터를 바탕으로 최적의 경로로 자동차를 운행함, 각각의 무인자동차들의 경로를 최적화하여 도로 위의 모든 차량은 최적의 경로로 운행할 수 있음

9.4

인공지능의 적용 분야



[표 9-5] 구글(Google)의 무인자동차 구조

단계	기능	핵심기술
데이터 수집	자동차의 센서에서 데이터 수집	사물인터넷(IoT)
저장 및 분석	지원 센터로 데이터 송출 및 분석	클라우드(Cloud), 빅데이터(BigData)
가치 창출	경로탐색, 자동 운행 서비스 제공	인공지능(AI), 딥러닝
최적화	최적 경로로 차량 운행	기술융합(convergence)

- ➔ 미국의 지능형 법률자문회사 로스인텔리전스(**ROSS Intelligence**)는 IBM의 인공지능 왓슨(Watson)을 기반으로 대화형 법률서비스(ROSS, 인공지능 변호사)를 제공하고 있음
- ➔ 단순히 키워드 검색 결과를 나열하는 기존 법률정보 검색시스템과는 달리 이용자가 일상의 대화체로 질문하면 질문과 연관성이 높은 법률적 답변과 함께 판례 등의 근거 자료를 제공함
- ➔ 유통과 물류 영역에서도 인공지능이 적극적으로 활용됨
- ➔ 아마존(Amazon)은 기존의 주문과 물류가 이원화된 시스템에서 주문, 재고, 유통을 통합한 주문이행센터를 고안하여 고객의 다양한 주문을 예측할 수 있는 인공지능 기반 예상배송시스템을 구축함

[표 9-6] 아마존의 예상배송시스템

단계	기능	핵심기술
데이터 수집	고객의 구매정보 데이터 수집	사물인터넷(IoT)
저장 및 분석	구매 패턴 분석	클라우드(Cloud), 빅데이터(BigData)
가치 창출	예상 구매 물품 예상	인공지능(AI)
최적화	유통과 물류 비용 감소	기술융합(convergence)

- ➔ 최근 인공지능은 예술의 영역에 도전하기 시작함
- ➔ 인공지능을 활용한 안면인식, 영상인식, 생체인식 기술이 크게 발전하여 분별의 정확도가 인간 능력의 이상으로 향상됨
- ➔ 구글과 마이크로소프트는 딥러닝 기술을 이용하여 대량의 영상으로부터 고양이 얼굴을 스스로 학습하는 인공지능을 개발함
- ➔ 인공지능(안면인식 기술)을 활용하여 마이크로소프트는 렘브란트 미술관과 공동으로 인공지능 프로젝트인 넥스트 렘브란트(Next Rembrandt)를 수행하고 있음
- ➔ 네덜란드의 대표적인 화가 렘브란트(Rembrandt)의 작품 300점 이상을 학습한 후에 그 화풍을 재현하여 렘브란트의 신작을 그림