

목 차

인공지능기반 챗봇 서비스 적용을 위한 선진사례 연구

2020년 1월

통 계 청

(박윤영, 문정민, 심현범)

□ 국외훈련 개요	1
□ 훈련기관 소개 및 요약	2
- 훈련기관 목적 및 방문 결과 -	
I. 서론	11
II. 본론	15
1. 인공지능이란	15
1. 1. 인공지능의 개념과 목표	
1. 2. 인공지능의 주요기술	
1.2.1. 학습	
1.2.2. 추론	
1.2.3. 인식	
1. 3. 인공지능의 기술 동향	
1.3.1. 기계학습	
1.3.2. 전이학습	
1.3.3. 강화학습	
2. 인공지능 기술을 활용한 챗봇	25
2. 1. 챗봇의 개념	
2. 2. 챗봇의 분류	
2.2.1. 동작방식에 따른 분류	
2.2.2. 정보교환방식에 따른 분류	
2.2.3. 답변생성방식에 따른 분류	
2. 3. 챗봇 서비스를 위한 주요 기술	
2.3.1. NLP(Natural Language Process)	
2.3.2. NLU(Natural Language Understanding)	
2. 4. 챗봇 보안에 관련하여	
3. 인공지능 챗봇서비스 활용 사례	31
3. 1. 국외 챗봇서비스 사례	
3. 2. 국외 챗봇서비스 사례	
III. 결론	37
[참 고]	39

국외훈련 개요

1. 훈련국 : 미국

2. 훈련기관명

- 미국 UMS (United Merchant Service)
- IEEE ICMLA Conference

3. 훈련분야 : 부서 맞춤형

4. 훈련기간 : 2019. 12. 11. ~ 21.

□ 훈련기관 소개

1. 미국 UMS

명칭	미국 UMS (United Merchant Service)		
소재지	미국 255 Route 17 South, Hackensack, NJ 07601		
홈페이지	www.unitedmerchant.com		
기관소개	<ul style="list-style-type: none"> ○ 1994년에 설립된 온, 오프라인 상의 결제 시스템을 운영하는 기관으로 자체 서버와 DB를 운영하고 있음 ○ 미국 전역의 방대한 양의 고객 DB를 보유하고 있으며, 위험 관리프로그램(risk monitoring program)을 운영·관리 ○ 자체 보안기술(Encryption)을 개발하여 패스워드 전송, 고객관리 등에서 발생할 수 있는 보안이슈 관리 ○ 고객의 편의에 맞춘 온라인(PC, 모바일 앱 등) UI, UX를 구축한 서비스 ○ 구현되어 있는 홈페이지기능의 차이를 확인하고, 이를 제공하기 위한 DB 구축방법에 대하여 토론 		
방문목적	<ul style="list-style-type: none"> ○ 통계 DB 구축, 활용방법 토의 <ul style="list-style-type: none"> - 구축 DB를 활용한 검색기능 토의 ○ PC, 모바일 서비스 UI, UX 토의 <ul style="list-style-type: none"> - 편의성 강화방안 토의 ○ 통계홈페이지 발전방안 토의 <ul style="list-style-type: none"> - 이용자 중심의 통계자료 제공을 위한 방안 토의 ○ 보안 및 개인정보 이슈 사례 등 학습 등 <ul style="list-style-type: none"> - DB활용시 발생할 수 있는 보안 및 개인정보 문제 해결방법 토의 		
교섭창구	Brenda Gracia Head of Risk Maagement United Merchant Services Inc.		
	전화 : +1-800-260-3388	FAX : none	E-mail :brenda@unitedmerc hant.com

2. IEEE(Institute of Electrical and Electronics Engineers)

명 칭	IEEE ICMLA Conference
소 재 지	미국 Boca Raton, Florida
홈 페이지	www.icmla-conference.org/icmla19
기관소개	<ul style="list-style-type: none"> ○ 2003년 이후 매년 머신러닝기술에 대한 컨퍼런스(ICMLA) 개최하여 현재 18회 운영 ○ 해당 컨퍼런스에서는 머신러닝과 딥러닝에 대한 전문가들의 수준높은 세션, 인터뷰 등이 진행 ○ 특히, 인공지능기반 챗봇서비스의 주요 기술인 머신러닝과 딥러닝(Machine and Deep Learning)에 대하여 주요 이슈 및 활용 사례에 대한 전문지식 습득
방문목적	<ul style="list-style-type: none"> ○ Machine Learning 학습 ○ Deep Learning 학습 ○ Machine, Deep Learning 사례 학습 ○ Text mining 이해 ○ Machine and Deep Learning in Cybersecurity and Privacy Issues 학습

□ 훈련 목적

- 국가통계포털(KOSIS) 인공지능기반 챗봇 서비스를 구축함으로써 정부의 공공데이터 확대 정책에 발맞추고 국가자산인 통계자료의 사회적 활용 가치를 높일 뿐만 아니라, 이용자들이 보다 쉽게 통계를 검색하고 이용할 수 있도록 함
 - 국가통계포털 이용건수는 매년 증가 추세
 - * 2015년 약 20백만건에서 2018년 55백만건으로 2.75배 가량 증가
 - 국가통계포털에서 제공하고 있는 통계표는 약 16만여개로 이용자가 원하는 통계를 손쉽게 찾아 활용할 수 있는 방안 마련이 필요
- 인공지능기술의 기본이 되는 머신러닝(Machine Learning), 딥러닝(Deep Learning) 기술의 현 상태를 파악하고 비전을 확인함으로써 선진기술 반영의 기회 획득
- 데이터 축적 및 분석을 통하여 인공지능기술의 자연어 처리 능력을 향상시키고, 효율적인 데이터 학습 방안에 대한 선진국의 모범사례와 전문기술자들의 노하우를 흡수하여 향후 국가통계포털 통계검색 기능(통합검색, 챗봇서비스)에 반영하고자 함
- 통계 이용자들의 국가통계포털(KOSIS) 이용에 관한 DB 구축 시 발생할 수 있는 위험(보안, 개인정보 이슈 등)에 대한 관리방안, 노하우 등을 습득

□ 훈련의 필요성

- 인공지능기술의 핵심인 머신러닝(Machine Learning), 딥러닝(Deep Learning) 주요내용 습득
 - 머신러닝은 기계학습으로 국가통계포털의 방대한 데이터를 시스템에 기억하게 하고 여러 상황들을 학습시킴
 - 딥러닝을 통하여 심화학습을 시켜 이용자의 질문에 대하여 스스로

- 생각하고 심화학습을 하여 적절한 응답을 이끌어 낼 수 있도록 함
- 머신러닝과 딥러닝을 통하여 학습된 데이터베이스를 기반으로 현재 보유하고 있는 16만 여개의 통계표를 국가통계포털 이용자들이 쉽게 찾아보고 활용할 수 있는 시스템 기반 마련
- IEEE에서 주관하는 ICMLA Conference에서는 이러한 기술 전문가들의 세션을 운영하고 인터뷰함으로 현재까지의 활용사례와 앞으로의 비전뿐만 아니라 취약점까지 해당기술의 모든 분야에 대한 지식을 제공함.
- 온라인 민원서비스가 증대에 따른 민원의 24시간 응답을 적절히 운영하기 위하여 선진기관의 통계DB 구축 방법과 활용방법에 대한 벤치마킹

○ 양방향 서비스 제공 필요

- 통계청의 국가통계포털은 통계생산부서에서 입력한 통계를 한곳에 모아서 보여주는 방식으로 서비스가 제공됨.
- 검색기능이 구현되어 있으나, 이미 입력되어있는 DB에 대한 자료만 검색이 되고 추가적인 옵션이나 학습기능 등은 구현되어 있지 않아 이용자들이 원하는 통계를 쉽게 찾지 못하는 상황 발생
- UMS의 온라인 서비스 전문가와 IEEE 세미나의 인공지능 전문가와의 자문을 통하여 단순 빈번한 통계표 검색기능에 학습데이터를 적용시켜 이용자와 제공자의 양방향 소통을 통한 챗봇 서비스 제공 방안 도출
- 해외 선진 시스템 운영 사례를 검토하여 국내 실정에 맞는 시스템 운영 체계에 대한 방안을 수립하는 데 있어 시사점을 얻을 필요

○ 온라인서비스 보안, 개인정보 이슈에 대한 대책

- IEEE주관으로 개최하는 ICMLA Conference에서 머신러닝과 딥러닝에 관한 보안 및 개인정보에 관한 이슈를 다루는 세션 참석
- 서비스를 운영하면서 발생하는 보안 취약점에 대하여 UMS의 선진 사례를 확인하고 이를 예방함과 동시에 이용 편의성을 높일 수 있는 방안을 선진 사례 학습을 통해 국내 적용 시사점 파악 필요

□ 훈련기관 방문 결과

○ UMS 방문결과

- 조직과 사업에 대한 개요

- UMS는 델라웨어를 기반으로 하는 UIPS LLC 지주사가 모기업임, UIPS는 UMSI 지주, UMSI, UPOS Holdings, United Pos, IMU를 관계회사로 두고 있음
- 마케팅 기관인 닐슨에 따르면 뉴저지에 기반을 둔 UMS는 상거래 기반 솔루션 업체 중 46위를 차지하고 있고, 판매 및 마케팅, 회계처리, 리스크관리 등 사업을 영위하고 있음
- UMS는 경쟁력있는 소비자 가격에 제공할 수 있는 통합솔루션을 개발하였고 하드웨어와 소프트웨어를 동시에 시장에 제공하고 있음
- 하드웨어로는 스캐너, 프린터, 윈도우 및 안드로이드 지원 POS, 현금인출기, 고객자동주문 KIOSK 등 다양한 결제 기기 등이 있음
- 소프트웨어로는 클라우드 기반 관리프로그램인 Data 분석기, 고용 관리, 고객관리, 디지털 마케팅 등이 가능한 소프트웨어를 개발 제공과 아울러 자영업자와 소상공인을 위한 고객맞춤형 솔루션도 개발하여 보급하고 있음
- 아울러 충성고객 관리를 위한 쿠폰.상품권 관리, 온라인 예약, 주문, 배달, 데이터분석에 이르는 부가서비스 프로그램도 제공하고 있음

- 데이터 수집 경로에 대한 상세한 설명

- 전통적인 단말기로는 15인치 화면용 터미널이 활용되고 있음
- 10~12인치 태블릿용 솔루션 탭도 있고, 소프트웨어와 부가서비스에는 재고관리, 고객관리, 고용관리 솔루션과 상품권, 주문 관리, 고객충성확보 솔루션이 있음
- 대표적인 사례로 'Feel Beauty'sms HW인 UP-5800I와 UP Retail Enterprise 솔루션을 통해 전미 30개 매장에서 Data를 관리
- 또 다른 사례로는 정부종합청사 편의점에 UP-5800(H/W) UP

Restaurant 솔루션을 통해 서비스를 제공하고 미국농무부, 국무부, 국회도서관에서도 활용함

- 글로벌 진출현황을 살펴보면 일본, 미국, 러시아, 이스라엘, 스페인, 영국, 그리스 등 약 30개 국가에 H/W와 S/W를 보급하고 있음
- 기존 H/W와 S/W 제공 판매를 통해 시장을 주도하고 있고 새로운 상품으로 21.5인치 자동터치스크린을 개발하여 시장에 출시하고 있음
- 단말기 혹은 전자상거래 채널을 통해 거래자료가 클라우드에 축적됨. 특히, 윈도우 혹은 안드로이드 기반 솔루션을 다 지원하는 통합 관리시스템을 통해 누적된 data를 주문, 예약, 배달 등 부가정보와 연계하여 분석보고서를 제공

○ IEEE 참석결과 및 발표 논문 동향

- 논문 동향 소개

- 전미전기전자학회의 제 18차 머신러닝과 응용에 대한 국제회의에서 4일간에 걸쳐 약 300편의 논문이 발표되었고 기계학습의 이론, 실제사례, 알고리즘 등 다양한 논문이 발표되었음
- (이미지처리) 이미지 인식에 대한 규제, 단일 이미지 최대화질 데이터 불호가도 추정 집중모델, 아키텍처 다중 부호화 및 비부호화 실시간 이미지 제고 등
- (딥러닝 알고리즘) 관련 물리입자 장단기 기억네트워크, 신경망을 활용한 공정성 중간감독 학습 레버리징 등
- (구조 모델링) 안식과 탐지를 주제로 원거리 센서와 기계학습을 활용한 이미지 탐색, 심화수렴신경망을 활용한 음성 및 안면부문에 기반한 양감정 인식모형
- (응용) 금융 및 마케팅 기계학습을 통해 심층신경망을 활용한 선물 시장 변동 예측, 실패예측을 통한 위험기반 동적 가격설정 등
- (텍스터 분석) 내재화 문자 학습에 대한 텍스터 연결망 형성에 기반한 비지시적인 주제
- ... 벡터 공간모형에서의 텍스트유사형, 짧은 text논평을 활용한 크

리켓운동선수들의 장단점 마이닝

- (컴퓨터 시각의 인식과 감지) 스마트홈 애플리케이션을 위한 심층잔여 u-net을 활용한 방사선 기반 목표탐지, 3층 합성신경망에 기반한 영상 오류 인식등
- (심층강화학습) 다중로봇임의 유형 형성, 스타크레프트 2 미니게임을 위한 계층화 모형
- 객체인식주제와 고나련하여 식물분류를 위한 최첨단 비주얼 시스템, 광업 디지털 감식을 위한 객체감지기반 심화학습
- 신경망 강건성 개선을 위한 환류학습, 자연어 군집내 강인성대표 학습과 분해
- (통제된 학습에프리케이션) 효율적인 공급자 분류에 대한 통제된 기계학습 접근, 논 버벌구 감지, 예측, 침투 분류화 형태분석
- (딥러닝 응용) 무선 다중네트워크이동 모형을 위한 분류자 기반 심층 신경망, 국제분류인식상표 접근에 대한 심층학습
- ... 효율적인 신경망 가지치기를 위한 활성화 및 가중에 관한 연계규제
- (SNS 및 이벤트마이닝) 소란스러운 고압적 공동체와 기부금 캠페인을 활용한 선거예측, 핵심성과지표에 영향을 끼치는 사건평가
- (불균형 자료) 브라질 도시에서 현대판 노예위기를 인식하기 위한 기계학습모형, 불균형계층 빅데이터 기준점과 딥러닝
- (자연어 처리) 외부보상 강화학습을 활용한 자연어 생성, 발음구별 부호 복원에서 입 출입 탐구 등이 발표
- (자율주행) 자율운전을 위한 자동 CNN 입출시스템, 외삽에 기반한 강화학습 모형을 활용한 이산선변동 결정
- (청각데이터에 관한 기계학습) 청각자료에 관한 기술어 집합을 활용하여 험백고래의 소리를 음악과 비음악으로 분류, 사전훈련된 합성신경망의 심층 혼합을 활용한 청각장면 분류

* 21세션에서는 기계학습 응용에서는 가 발표됨

□ 훈련결과보고서 요약서

훈련자	박윤영 문정민 심현범	직 급	통계사무관 행정주무관 통계주무관
소속	통계청		
훈련국	미국	훈련기간	2019.12.11. ~ 2019.12.21.
훈련기관	UMS, IEEE	훈련구분	부서맞춤형
훈련목적	통계서비스 확대	보고서매수	39매
내용요약	<p>현대사회는 4차산업의 시대라고들 한다. 스마트폰이 일반적으로 보급되어 언제 어디서든 휴대폰으로 인터넷을 할 수 있고, 많은 정보를 얻고 서비스 받을 수 있다. 이러한 기술은 더욱 발전하고 있다.</p> <p>1956년 최초로 인공지능(AI, Artificial Intelligence)의 용어가 John McCarthy에 의하여 “지능형 기계, 특히 지능형 컴퓨터 프로그래밍을 만드는 과학 및 공학”으로 정의되고 사용된 이후로 현재까지 많은 연구가 이루어지고 있고, 서비스 되고 있다.</p> <p>인공지능 기술 중 하나인 기술학습(Machine Learning), 딥러닝(Deep Learning) 등을 활용하여 마치 컴퓨터가 인간과 비슷하게 생각하고 행동할 수 있게 된다. 딥러닝 기술을 활용한 비 지도학습(Unsupervised learning)이 가능해지고 기계가 방대한 양의 데이터를 축적하고 학습하게 되면서 점점 정확하고 효율적인 값을 출력하는 것이 가능해진다. 또한, 인간의 언어를 자연어처리(NLP, Natural Language Process)를 통하여 컴퓨터 언어로 변환하여 이해(NLU,</p>		

<p>Natural Language Understand)하고 응답한다. 질문에 대한 일반적인 답만을 제공하는 것이 아니라 정보교환의 형식의 대화도 가능하다. 처음에는 원활한 대화를 하지 못할지 모르겠지만 시간이 갈수록 학습하고 대화를 익히게 되면서 부터는 마치 사람과 대화하는 것과 비슷한 대화가 가능해질 것이다.</p> <p>인공지능 챗봇 서비스는 각종 커뮤니티, 기관서비스, 기업서비스를 이용할 때 편리함을 제공해준다. 제공되는 서비스가 많아질수록 홈페이지는 복잡해지고, 서비스 이용방법 또한 복잡해진다. 필요한 서비스를 찾는 데는 점점 더 시간이 많이 소요되고 결국 찾지 못할 때에는 상담원의 안내가 필요하게 된다. 이 때, 챗봇 서비스를 이용한다면 언제 어디서든 몇 번의 채팅, 키워드 검색만으로 원하는 서비스를 찾을 수 있다. 여기에 인공지능 기술이 더해져 사람과 대화하는 것처럼 입력을 하더라도 자연어 처리 기술 등을 사용한 서비스가 가능하다. 딥러닝 기술을 활용하여 챗봇 서비스는 학습을 하게 되면서 시간이 가고 데이터가 축적될수록 더욱 똑똑해져서 질문을 하고 응답을 출력하는데 더욱 디테일한 정보제공이 가능해진다.</p> <p>하지만 인공지능기술이 발전하고 정보의 양이 많아질수록 축적되는 개인 데이터도 많아진다. 그에 따른 유출 위험도 많아지고 보안의 중요성도 커진다. 보안을 유지하기 위해서는 축적되는 데이터에 필요 없는 자료를 삭제하는 정제작업, 엄격한 보안정책, 보안프로그램 등 다양한 부분에서 관리를 하여야 하며 유의해야 한다.</p> <p>기술적인 부분과 보안위험성 제거 등 개인이 정보를 안심하고 이용할 수 있도록 많은 노력이 필요하다.</p>

I. 서론

현대사회는 일상생활부터 업무적인 측면까지 많은 부분이 자동화 되어 이루어진다. 몇 년 전만 하더라도 주민 센터를 방문하여야만 발급받을 수 있었던 주민등록등본·초본 등과 같은 서류가 이제는 컴퓨터와 프린터만 있으면 얼마든지 무료로 발급받을 수 있다. 어떠한 것이든 궁금하거나 필요한 것이 있으면 스마트폰을 통하여 관련된 회사 또는 부서로 전화하여 문의할 수 있고, 포털사이트 등을 통하여 검색해 볼 수 있다. 그야말로 정보화의 시대이다.

흔히들 현재 4차 산업혁명의 진입시대에 살고 있다고들 한다. 하지만 많은 제공하는 정보가 많아질수록 이용방법은 복잡해지고 원하는 자료는 찾기 어려워지기 마련이다. 매년 서비스를 개선하고 편의를 위한 노력을 한다고 하더라도 제공하는 정보의 양은 점점 늘어나 복잡해 질 수밖에 없다. 이러한 상황을 해결하기 위한 대책으로 인공지능(AI)기반의 정보제공 서비스가 부각되고 있다.

“지능형 기계, 특히 지능형 컴퓨터 프로그래밍을 만드는 과학 및 공학” 미국의 저명한 컴퓨터 공학자인 John McCarthy가 1956년 최초로 인공 지능(AI, Artificial Intelligence)라는 용어를 다트머스 회의의 회의 명으로 지정하면서 정의하였다.

그리고 인공지능의 선구자인 마빈 민스키 교수는 “인간은 생각하는 기계다” 라고 하였다. 그는 뇌 신경망을 모방해 컴퓨터의 논리회로를 만들면 컴퓨터도 지능을 가질 수 있다고 생각했다. 인간의 뇌를 이루는 수많은 신경세포인 뉴런은 자극을 받았을 때 전기를 발생시켜 다른 세포로 정보를 전달할 뿐, 각각의 세포 자체가 기능을 갖고 있지는 않다고 하였다. 이런 생각은 논리 회로를 어떻게 연결하느냐에 따라 인간처럼 생각하는 기계를 만들 수 있다는 확신으로 이어졌다. 그가 발표한 ‘지적 활동의 프레임워크(A Framework For Representation Knowledge)에서는 인간의 지식활동, 지적활동을 외부로부터 입력된 데이터와 내부프레임의 상호작용으로 보는 연구를 하였다.

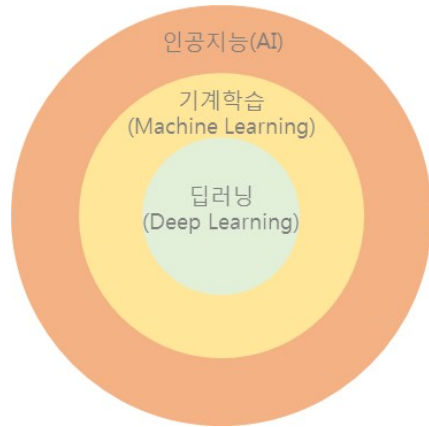
인공지능은 간단하게

- 컴퓨터에서 지능적인 행동을 시뮬레이션 하는 컴퓨터 과학 분야
 - 지능적인 인간의 행동을 모방하는 기계의 능력
 - 일반적으로 사람의 지능을 필요로 하는 작업을 수행할 수 있는 컴퓨터 시스템
- 등으로 정의 할 수 있다.

즉, 사람의 지능을 통해 하던 행동을 컴퓨터 시스템을 활용하여 분석하고 학습하여 판단하는 기술이라고 생각할 수 있다.

인공지능 관련 기술은 컴퓨터 과학 분야의 여러 영역과 서로 연계되어 있으며, 여전히 연구하고 발전하는 단계에 있다. 인공지능 관련 기술은 여러 가지가 있다. 경험적 데이터를 통해 컴퓨터 스스로 새로운 지식과 능력을 개발하는 기계학습(Machine Learning), 인간의 두뇌 처리기능을 형상화한 기술인 인공 신경망(Neural Network), 기존 다층 구조의 인공 신경망 학습의 문제점을 개선하여 인공신경망이 적절히 학습할 수 있도록 돕는 딥러닝(Deep Learning), 인간의 언어나 사고가 가진 애매한 정도를 수학적으로 다루고자 하는 논리체계인 퍼지논리(Fuzzy Logic)가 대표적이다. 그 외에도 패턴인식(Pattern Recognition), 자연어처리(Natural Language Processing), 가상현실(Virtual Reality) 등이 있다. 이러한 분류는 아직까지 정확한 체계가 잡혀지지 않고 있지만 전 세계적으로 기계학습(Machine Learning), 인공신경망(Neural Network), 딥러닝(Deep Learning)을 기반으로 한 인공지능 기술에 대한 연구는 활발히 진행되고 있다.

이들 간의 관계를 간단하게 표현해 본다면 다음과 같다.



인공지능 기술은 여러 국가, 기관이 다양한 분야에서 활용하고 있다.

싱가포르는 국가인공지능전략을 가동('19.11.)하여 “교통 및 물류”, “헬스케어”, “안전 및 보안”, “스마트시티 및 부동산”, “교육”의 5분야를 선정하여 프로젝트를 시작한다고 선언하였다.

미국은 인공지능으로 라스베이거스의 교통난을 해소하기 위한 교통시스템 개발에 착수하였다.

국제농업연구협의그룹(CGIAR, Consultative Group on International Agricultural Research)은 농업분야에서 빅데이터와 인공지능을 활용해 기아 및 미래의 식량위기에 대비하기 위하여 노력하고 있다. 농업인들이 데이터 패턴분석을 통하여 농작물의 식재 장소와 시기, 비용 절감 등에서 인공지능을 활용한다는 방식이다.

우리가 흔히 알 수 있는 인터넷에서도 쉽게 인공지능을 활용한 기술을 접할 수 있다.

그 중 가장 가까워서 쉽게 접할 수 있는 기능 중 하나는 챗봇(Chat-Bot)이다. 우리가 흔히들 말하는 챗봇은 대화형 인공지능 기술에 해당한다. 각종 커뮤니티, 홈페이지 등을 이용할 때 더욱 편리하게 원하는

기능을 찾고, 활용할 수 있도록 도와주는 기능이다. IT 기술이 발달하고 언제 어디서든 인터넷 사용이 가능함에 따라 서비스 이용자들의 문의는 점점 늘어나고 있다. 이러한 문의를 사람이 직접 응대하는 것이 아닌 학습된 기계가 원하는 자료를 정확하고 빠르게 찾아주는 것이다. 학습 데이터가 축적되고, 이를 수정하고 다시 학습하는 등의 반복적인 활동이 계속되면 될수록 더욱 정확한 자료를 찾아줄 수 있게 된다. 이에 이용자는 담당자와의 직접적인 연결을 기다리지 않아도 질문 한두개 만으로 원하는 서비스를 찾아 활용할 수 있게 되는 것이다.

통계청에서 운영하고 있는 국가통계포털은 국내·외의 통계를 한곳에 모아서 한번에 확인할 수 있도록 서비스하고 있다. 2007년에 서비스를 시작하여 2019년 12월 기준 388개 기관에서 1,226가지의 통계를 서비스하고 있으며, 제공되고 있는 전체 통계표의 수는 약 16만여 개에 해당한다. 이것은 단지 통계표로 분류한 수치일 뿐이고, 통계표안의 항목까지 포함한다면 셀 수 없을 정도의 정보를 제공한다. 그 중에서 이용자는 원하는 통계항목을 검색하고, 통계표를 확인하여 활용할 수 있다. 하지만 수많은 통계표와 항목에서 원하는 것을 찾아내기란 쉽지 않다. 단순한 민원의 범위를 벗어나, 통계표의 구성, 용어, 년도, 지역 등의 많은 부분이 유기적으로 연결되어 있기 때문이다. 때문에 통계 검색에 어려움을 겪고 있다는 민원이 아주 많이 발생하고 있다. 많은 양의 데이터를 가지고 있기 때문에 이러한 민원을 담당자가 직접 처리하여 알려주는데도 시간이 많이 소요된다. 따라서 인공지능기술을 활용하여 검색기능을 강화하고, 자연어 처리에 대한 학습, 통계 명, 항목 등에 대한 DB축적 등을 통하여 시스템화가 이루어진다면 이용자뿐만 아니라 통계정보 제공자까지 많은 부분에서 편리할 것으로 기대된다.

II. 본론

1. 인공지능이란

1.1. 인공지능의 개념과 목표

인공지능은 여러 차례의 기술적 변화를 통해 단순 문제풀이 위주나 독립적인 분야로서의 발전이 아니라 지능의 실제적 구현을 목표로 발전되고 있으며, 특히 ICT 기술의 발전으로 다양한 분야로의 융합으로 발전하고 개발되고 있다.

인공지능은 여러 학자들에 의해 개념이 정의되고 있다. 1956년 미국 다트머스 대학에서 열린 회의에 존 매카시 박사가 다스 학회에서 처음으로 인공지능이라는 용어를 창안하였으며, "기계를 인간 행동의 지식에서와 같이 행동하게 만드는 것"이라는 인공지능의 정의를 내렸다.

또한, 영국의 수학자 앨런 튜링(수학자, 암호학자, 논리학자이자 컴퓨터 과학의 선구적인 인물)은 컴퓨터가 사람처럼 생각할 수 있다고 제시하여, 대화를 나누고 상대방이 컴퓨터인지 사람인지 구별할 수 없다면 그 컴퓨터는 사고할 수 있는 것이라고 간주해야 한다고 주장하였다.

인공지능의 목표는 다음과 같이 크게 4가지로 분류할 수 있다.

<인공지능의 정의 및 목표>

인간의 사고작용 (thinking)	인간과 같은 사고 시스템 (systems that think like humans)	합리적 사고 시스템 (systems that think rationally)
	인간과 같은 행동시스템 (systems that act like humans)	합리적 행동 시스템 (systems that act rationally)
행동 (behavior)		
	이론적(ideal)	합리적(rational)

먼저 인간과 같은 사고 시스템은 이론적으로 인간처럼 생각하는 기계를 만들려면 우선 인간의 사고 작용을 연구해야 하고 이로부터 그럴듯한 가설이 정립되면 프로그램을 통해 실현할 수 있다는 것을 주장하는 분야이다.

둘째, 합리적 사고시스템은 인공지능의 목표를 이론적이 아닌 합리적으로 사고하는 시스템이라고 정의하는 분야이다.

셋째, 인간과 같은 행동시스템은 컴퓨터는 인간에게 질문함으로써 그의 테스트 결과를 여러 경로를 통해 보내주게 된다는 것을 주장하는 분야이다.

마지막으로 합리적인 행동시스템은 주어진 확률 정도가 있을 때 어떤 목표를 달성하기 위해 행동하는 것을 의미한다.

1.2. 인공지능의 주요 기술

인공지능의 주요 기술은 크게 “학습, 추론, 인식, 행동” 등으로 분류할 수 있다.

1.2.1. 학습

인공지능의 주요기술인 “학습”을 이해하기 위해 인간의 뇌를 단순화하여 구현한 것이 신경회로망이며, 이것이 업그레이드 된 기술이 최근의 딥러닝 분야이다.

신경회로망에서 학습 개념은 매우 중요한데, 한 처리기의 지식 변화는 인접된 다른 처리기에도 변형을 주며, 기초의 연결 강도수정으로 이루어진다. 이러한 연결강도는 경험적으로 변형되며, 연결강도의 변화를 학습 규칙이라 한다.

(1) 지도학습(supervised learning) : 입력신호에 대한 정답 출력을 주는 학습 방식

이름에서 알 수 있듯이 컴퓨터에게 정답(Label)이 무엇인지 알려주면서 컴퓨터를 학습시키는 방법이다

예) $3x2=6$, $7x8=56$ 등을 학습시킨 후 $9x3=?$ 등의 주어진 문제를 해결하는 방식
: data : $3x6$, $7x8$ label : 27(정답)
⇒ 정확한 input과 output이 존재

· 분류(Classification)

주어진 데이터를 정해진 카테고리에 따라 분류하여 학습

예) 스팸메일과 일반메일로 분류 ⇒ 예/아니오로 구분 : Binary Classification
 동물(강아지, 고양이, 햄스터 등)
 ⇒ 예/아니오로 구분 안됨 : Multi-label Classification

· 회기(regression)

어떤 데이터의 특징(feature)을 토대로 값을 예측하는 학습방식으로 결과 값은 실수를 가질 수 있으며, 주로 어떤 패턴이나 트렌드, 경향을 예측할 때 사용한다.

예) Q : 30대 남성의 평균 키는 얼마 정도야?
 A : 30대 남성의 평균 키는 00cm입니다.

(2) 비지도학습(unsupervised learning)

평가기준은 있으나 일일이 교사 신호를 주지 않는 학습 방식으로 지도 학습과는 달리 정답을 알려주지 않고 비슷한 데이터를 군집화 하여 미래를 예측하는 학습방법이다. 라벨링(Labeling) 되어있지 않은 데이터로부터 패턴이나 형태를 찾아 학습하는 방식으로 복잡하고 어려운 학습방식이다. 지도 학습에서 적절한 특징(feature)을 찾아내기 위한 전처리 방법으로 비지도 학습을 사용하기도 한다.

· 군집(clustering)

예) 고양이, 강아지, 닭, 오리를 비지도 학습으로 분류 ⇒ 특징별로 분류
 - 다리가 4개 : 강아지, 고양이
 - 다리가 2개 : 닭, 오리
 ⇒ 비지도 학습은 예측 등이 아닌 데이터가 어떻게 구성되었는지 밝히는데 주로 사용(Grouping Algorithm)

1.2.2. 추론

정보에 대한 가정과 전제로부터 결론(지식)을 이끌어 내거나 도출해내는 기술이며, 개별적 정보를 이해는 단계를 넘어 정보 간 복잡한 관계를 파악하여 표현하는 기술이다.

1.2.3. 인식

인식은 학습을 바탕으로 새로운 자료나 불확실한 자료가 주어졌을 때 추론을 통해 알아차리는 과정이며, 그 결과를 표출하는 과정까지 포함한다. 인공지능에서는 다양한 인식 기술들이 있으며, 기본적으로 인식을 위한 단위를 패턴(pattern)이라고 하고, 하나의 패턴은 하나의 unit이나 개념을 표현한다. 대표적인 패턴으로는 글자인식, 영상인식, 음성인식, 개인 성향 인식, 상황인식, 위치인식 등이며 미래예측, 결과예측 등과 같은 패턴도 인식의 범위에 포함시킬 수 있다.

<인공지능 SW 기술의 범위>

유형	과학기술정보 통신부 (2018.05.)	국내 기관 (2018.02.)	해외기관	본 브리프 기술 범위
학습	머신러닝	학습 및 추론	Machine Learning	기계 학습
추론	추론·지식표현		Machine reasoning	지식추론
인지	시각지능	시각이해	Computer Vision	시각지능
	언어지능	언어이해	Natural Language	언어지능
	청각지능			
응용	상황이해	복합지능	-	-
	지능형 에이전트	기타 인공지능		
	행동·협업지능			

1.3. 인공지능의 기술 동향

과거 오랜 시간 소요되었던 기계학습 과정은 컴퓨터 성능, 알고리즘 진화 등을 통해 단기간 처리가 가능해졌으며, 방대한 데이터를 통해 기계는 마치 실제 세상 속에서 인간처럼 정보를 인지하고 학습해 지식으로 발전시켜 나가기 시작하였다.



1.3.1 기계학습(Machine Learning)

인공지능 기술의 핵심 기반기술인 기계학습은 데이터 기반의 효율적인 학습을 위한 새로운 기술들이 등장하고 있다.

인공지능의 목표는 “사람처럼 생각하고 행동하는 기계/컴퓨터”를 만드는 것이라고 볼 수 있다. 인공지능을 구현하고자 할 때, 지식과 규칙에서 시작해서 다양한 응용을 추구하는 top-down 방식과, 실시간의 데이터에서

시작해서 특정 응용문제를 해결하고 나아가 지식과 규칙의 발견까지도 기대하는 bottom-up 방식의 두 방향이 가능하다. 머신러닝(Machine Learning)은 bottom-up 방식의 대표적인 기술이고, 딥러닝(Deep Learning)은 머신러닝의 최신 기술/트렌드이다.

또한, 소량의 데이터에도 예측 정확도를 높이고 인간의 개입을 최소화하여 스스로 학습하는 등의 새로운 기계학습 방법들도 등장하고 있다.

- * 경영컨설팅기관인 맥킨지사는 기계학습 기술의 발전 방향과 적용 가능성을 기준으로 기술의 트렌드를 분석하였으며 그 중 가장 최신의 기술이며 적용 가능성이 높은 기술로 딥러닝, 전이학습, 강화학습이라고 언급
- * 머신러닝 절차 : [데이터 수집 및 정리] - [특징 추출] - [머신러닝 알고리즘 등을 적용하여 학습 수행] - [학습한 모델을 적용]

(1) 딥러닝(Deep Learning)

인간의 뇌 신경망 구조를 모방한 딥러닝은 다층의 레이어를 통해 데이터를 학습하는 방법이다. 딥러닝 과정에서는 ‘특징추출’ 과정도 학습 과정에 포함을 하여 데이터만 보고도 문제를 표현하고 푸는 데 좋은 특징을 학습을 통해 찾는다.

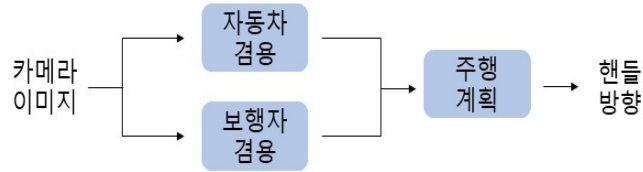
가) 종단학습(end-to-end learning)

: 학습에 필요한 데이터를 거의 원본 그대로 입력하여 지정한 형태의 출력이 산출되도록 학습하는 과정

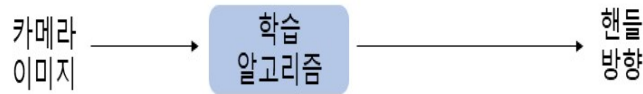
딥러닝을 이용해 지도학습, 비지도 학습을 할 때 주변에서 구할 수 있는 형태의 데이터(예) 알파벳, 사진픽셀파일 등))를 그대로 알고리즘의 입력 값으로 사용한다. 이러한 표현력 학습(representation learning)을 통해 입력과 출력 사이의 전 과정을 기계가 학습을 통해 자동으로 구축하는 학습해법이 등장하고 있다.

예) 자동차 자율주행

- 전통적인 방식 : 카메라 이미지에서 자동차를 검출하는 모듈과 보행자를 검출하는 모듈을 따로 구성한 후 두 정보를 취합하여 주행 계획을 세우고 최종적으로 핸들 방향을 조정



- 딥러닝 : 카메라 이미지를 입력하여 바로 핸들방향을 조정



* 대규모 데이터가 있어야만 구현이 가능하다.

나) 알고리즘

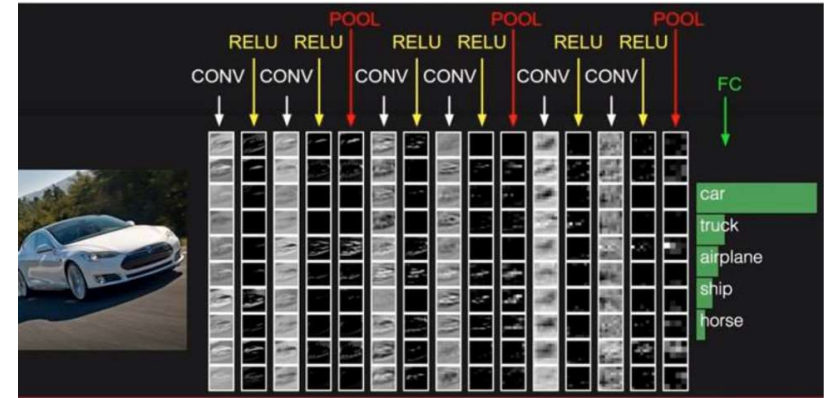
딥러닝은 시각지능, 언어지능 등에 특화된 학습 알고리즘이 지속적으로 개발되고 있고, CNN, RNN등의 알고리즘 등장으로 기계가 데이터를 통해 자신만의 규칙을 생성하여 정보를 학습하는 시대로 발전하였다.

i) 콘볼루션망(CNN, Convolutional Neural Network)

1955년 LeCun과 Bengio가 처음 발표하였고, 딥러닝의 기법 중에서 음성 인식 및 이미지 인식에 탁월한 성능을 보이는 알고리즘이다. 시각 피질의 신경세포들은 물체의 방향과 장소가 바뀌어도 별 문제 없이 인식할 수 있는 것을 생각하면 이해가 쉽다.

* CNN은 특징(Feature)을 추출하는 콘볼루션(Convolution)층과 정보를 압축하는 풀링층(Pooling)으로 이루어져 있다. 입력층과 출력층 사이에 콘볼루션층과 풀링층이 반복되며 결과값을 출력한다. 예를 들면 자동차 이미지 스캔을 반복작업하다가 바퀴 모양을 스캔하게 되면 자동차일 확률이 높으므로 높은 값을 출력하게 되는 방식이다.

<자동차 이미지를 스캔하는 과정>



* 출처 : brunch(<https://brunch.co.kr/@linecard/323>)

ii) 순환신경망(RNN, Recurrent Neural Network)

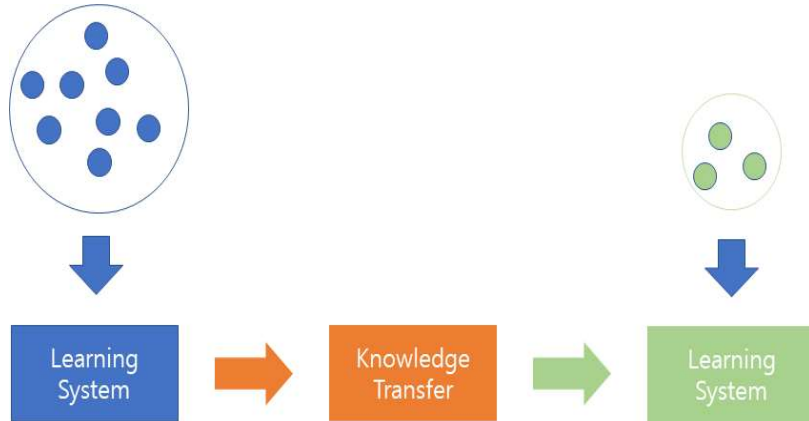
시계열 데이터와 같이 시간의 흐름에 따라 변화하는 데이터를 학습하기 위한 인공신경망이다.

* 과거의 출력 데이터를 재귀적으로 참조하게 된다. 즉 자기 자신을 참조하여 현재 결과가 이전결과와 연관성을 가지게 된다. 예를 들면 글자나 문자를 완성해주는 알고리즘을 대표적으로 들 수 있다. 사용자가 '동해물과'라고 입력하면 뒤에 '백두산이'가 이어질 확률이 높으므로 연관성 있는 '백두산이'를 출력하게 된다. 또한 음성인식 및 번역에도 많이 사용되고 있다.

1.3.2. 전이학습

특정 분야에서 학습된 신경망의 일부 능력을 유사하거나 전혀 새로운 분야에서 사용되는 신경망의 학습에 이용하는 것을 의미한다. 기존의 기계 학습방법은 훈련 데이터를 새롭게 수집하고 모델을 새로 만들어야 하는 과정이 필요하였으나, 전이학습은 학습 완료된 모델을 유사 분야에 전이(Transfer)하여 학습 시키는 기술로 적은 데이터에도 학습을 빠르게 하고 예측의 정확도를 높이는 방법이다.

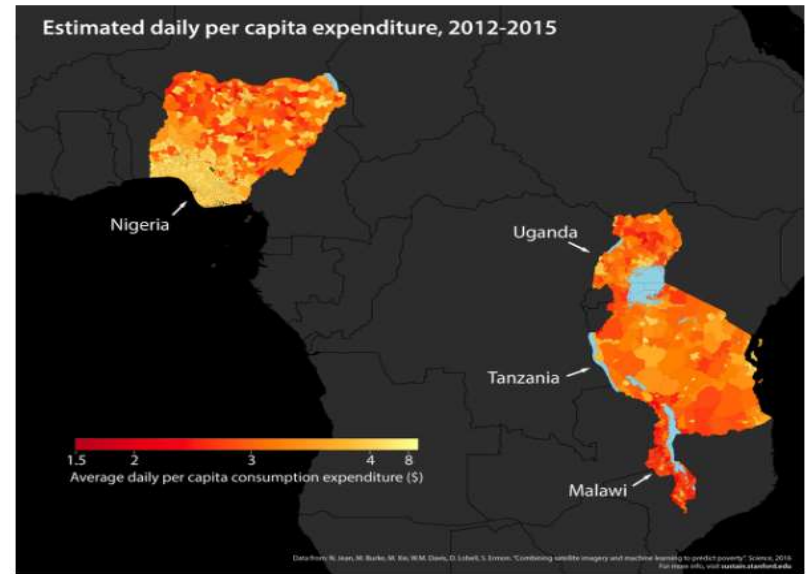
<전이학습>



전이 학습은 특정 환경에서 만들어진 AI알고리즘을 다른 비슷한 분야에 적용하는 것으로, 현재 많은 연구가 이루어지고 있다. 쉽게 예를 들어보면 사과 깎는 법을 익힌 AI에게 배를 깎게 시키는 것, 체스를 익힌 AI에게 장기를 두게 하거나 비를 예측하는 AI에게 눈을 예측하게 하는 것을 들 수 있다.

- 스탠퍼드대 연구진은 풍부한 야간 조명 데이터를 활용한 전이 학습 모델과 고해상도의 주간 위성사진을 결합하여 도로, 도시지역, 수로, 농지 등 빈곤과 연관할 수 있는 개체를 식별하는 알고리즘을 고안하였다. 이 알고리즘을 통해 AI는 개체들의 특징을 학습하여 빈곤 지역에서 발생할 야간 조명을 예측하고, 낮과 밤, 2개의 데이터 세트를 비교 하는 것으로 빈곤 지역을 이전보다 세부적으로 파악할 수 있게 되었다.

<빈곤 지도>



* 출처 : <http://sustain.stanford.edu/predicting-poverty>

1.3.3. 강화학습

행동심리학에서 영감을 받았으며, 어떤 환경 안에서 정의된 에이전트가 현재의 상태를 인식하여, 선택 가능한 행동들 중 보상을 최대화하는 행동 혹은 행동 순서를 선택하는 방법이다. 이러한 강화학습 문제를 공식화하여 그것을 추론 하는데 가장 일반적인 방식은 마르코프 의사결정과정이다.

(1) 마르코프 의사결정과정(MDP, Markov Decision Process)

마르코프 결정 과정은 어떤 상태 s 에 존재한다. 의사결정자는 해당 상태 s 에서 어떤 행동 a 를 취할 수 있으며, 다음 시점에서 마르코프 결정 과정은 확률적으로 새로운 상태 s' 로 전이한다. 이 때 의사결정자는 상태 전이에 해당하는 보상 $R_a(s, s')$ 을 받는다. 기존의 상태 s 에서 새로운 상태 s' 로

전이하는 확률은 의사결정자의 행동에 영향을 받는다. 즉, 전이 확률 함수는 $P_a(s, s')$ 와 같이 주어진다. 따라서, 다음 상태 s' 는 현재 상태 s 와 의사결정자의 행동 a 에만 영향을 받으며 이전의 모든 상태와는 확률적으로 독립적이므로 작용하게 되는 방식이다.

(2) Q-러닝

무작위 행동빈도(Exploration Rate)을 낮추어 가면서 세상이 주는 보상에 대한 경험을 쌓아 세상의 다양한 모습(State)에 대응할 수 있는 가장 올바른 행동양식을 찾아가는 학습방법이다. 이는 마치 사람이 태어나서 세상을 배워 나가고 경험을 활용하는 방식으로 처음에는 무작위로 행동하여 경험을 쌓고 나중에는 경험을 바탕으로 지혜롭게 행동하게 되는 알고리즘이다.

2. 인공지능기술을 활용한 챗봇

2.1. 챗봇의 개념

메신저 서비스는 최근 사용자수가 급증하면서 웹기반 SNS 사용자수를 2015년부터 추월하기 시작하였으며 이로 인해 채팅이 중요한 비즈니스 창구로 떠오르게 되었다. 또한 인공지능의 발전으로 자연어 처리 기법(Natural language processing)에서 자연어 이해기법(Natural language understanding), 자연어 생성기법(Natural language generation)으로 발전하게 되면서 챗봇이 재조명되고, 많은 IT 기업들이 챗봇 관련 플랫폼을 개발 및 발표하고 있고 공공기관에서도 도입을 발표하거나 검토 중에 있다.

챗봇이란 사람과의 문자 대화를 통해 질문에 알맞은 답이나 각종 연관 정보를 제공하는 '인공지능(AI) 기반의 커뮤니케이션 소프트웨어를 지칭한다. 챗봇은 사람들이 필요로 하는 서비스와 데이터를 적시에 찾아주는 등 수많은 기업과 개인을 연결해주는 연결고리 역할을 한다.

최초의 챗봇은 1966년도에 개발되었을 정도로 아주 오래된 기술이나 인공지능 기술의 발전으로 머신러닝 또는 딥러닝 기술이 챗봇에 적용되어 더 복잡한 일을 수행하는 것이 가능해졌다. 또한, 스마트폰의 보급과 모바일 메신저 사용자가 급증하여 스마트폰 사용자는 언제 어디서든 모바일 웹, 메신저 앱 등을 통하여 익숙한 환경에서 챗봇 서비스를 활용할 수 있다.

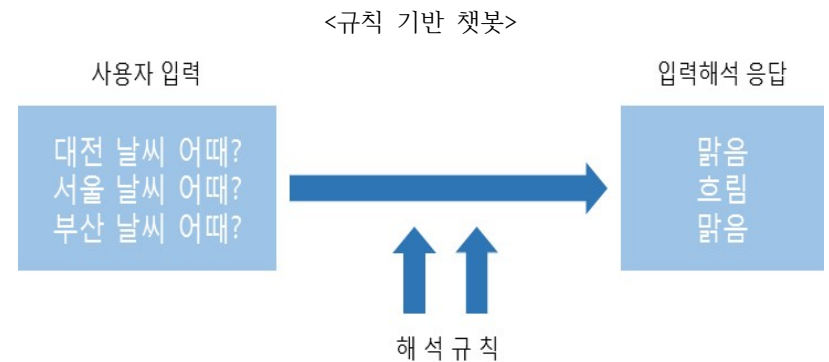
2.2. 챗봇의 분류

챗봇은 동작방식에 따라 규칙기반 챗봇과 기계학습기반 챗봇, 정보교환방식에 따라 일회성 질의응답 챗봇과 연속대화형 챗봇, 답변생성방식에 따라 검색모델 챗봇과 생성모델 챗봇으로 분류할 수 있다.

2.2.1. 동작방식에 따른 분류

(1) 규칙 기반 챗봇

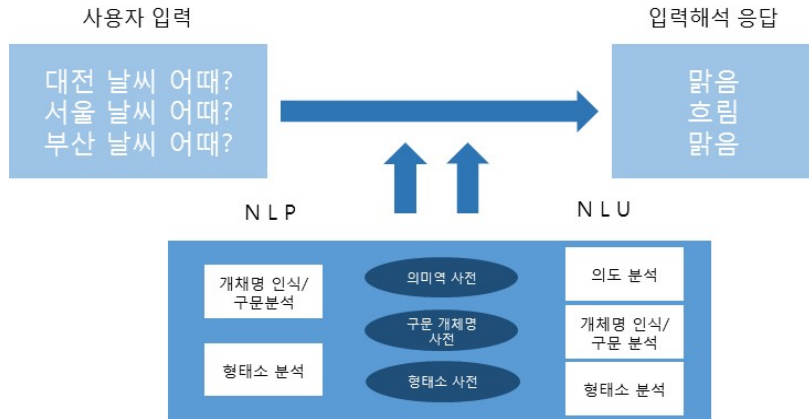
미리 정의된 규칙을 기반으로 챗봇이 동작하게 되어있어는 챗봇을 말한다. 사용자의 입력을 해석하는 해석규칙, 사용자의 입력에 반응하는 반응규칙 또는 응답을 생성하는 규칙을 포함하고 있다. 규칙기반 챗봇은 대량의 데이터가 필요하지 않으며 상대적으로 구현이 쉽지만 높은 품질의 대화 서비스가 제공되기 위해서는 규칙의 정의에 많은 시간과 인력이 소요될 수 있다.



(2) 기계학습 기반 챗봇

자연어를 이해하고 적절한 답변을 생성하거나 선택하기 위해 기계학습 알고리즘을 사용하는 챗봇이다. NLP(Natural Language Process)모듈, NLU(Natural Language Understanding) 모듈로 사용자가 입력한 자연어를 다양한 언어학적 자원을 활용해 분해하고, 분해된 자연어의 의미를 파악하여 최종 분류 모델 또는 생성모델로 전달한다. NLG(Natural Language Generation) 모듈이 결과물을 사용하여 새로운 답변을 생성한다. 기계학습 기반 챗봇은 높은 수준의 대화가 가능하지만 그만큼 높은 수준의 자연어 처리기술과 많은 언어학적 자원 및 데이터를 필요로 한다.

<기계학습 기반 챗봇>



2.2.2. 정보교환 방식에 따른 분류

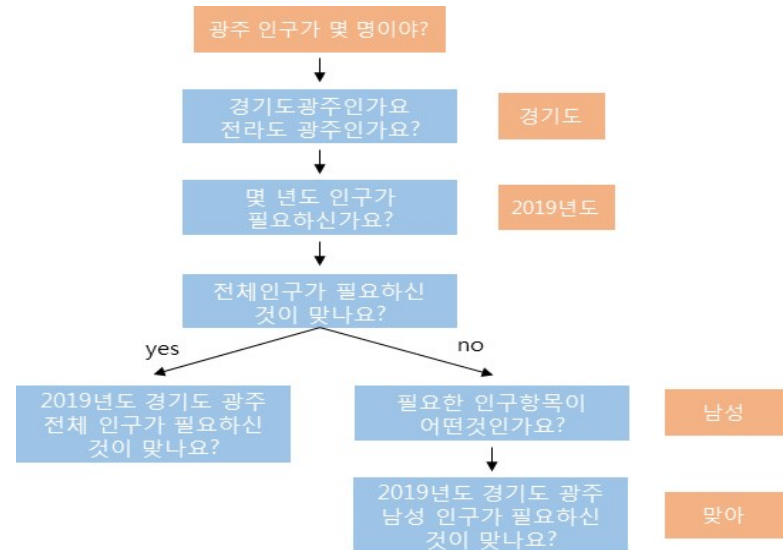
(1) 일회성 질의 응답 챗봇

대화관리 기능을 필요로 하지 않으며 질문과 관련성이 가장 높은 답변을 제시하는 검색엔진과 유사한 형태로 작동한다 자주 묻는 질문에 대한 정해진 답변을 제시하는 FAQ 챗봇의 형태로 1회성 질의응답 형태로 설계된다.

(2) 연속 대화형 챗봇

대화의 시작과 종료가 있으며 사용자로부터 특정한 목적을 달성하기 위해 정보를 수집하고, 수집된 정보를 바탕으로 최종 응답을 하면 대화가 종료된다.

<연속 대화형 챗봇>



챗봇이 모든 질문에 대한 응답이 불가능 하므로 대부분 특정 대화의 주제에 따라 구조화된 대화의 형식(시나리오)이 정해져 있다.

2.2.3. 답변생성방식에 따른 분류

(1) 검색기반 챗봇

사용자의 질문에 따라 기존 지식 베이스에 저장된 답변 중 가장 적절한 답변을 선택하여 제시하는 방식으로 작동하는 챗봇이다.

(2) 생성모델 챗봇

답변을 생성하는 자유도가 큰 방식으로 문법적인 오류를 범하거나 일반적이지 않은 답변을 생성하는 경향이 있다. 생성모델 챗봇은 굉장히 방대한 데이터를 필요로 한다.

2. 3. 챗봇 서비스를 위한 주요 기술

2.3.1. NLP(Natural Language Process)

자연 언어 처리(Natural Language Processing, 이하 NLP)는 컴퓨터와 인간 언어 사이의 상호 작용하는 기술로 인공지능의 핵심 기능 중 하나이다. NLP 기술은 기계번역, 대화체 질의응답 시스템 대화시스템, 정보검색, 말뭉치 구축, 시맨틱웹,텍, 딥러닝, 그리고 빅데이터 분석 분야뿐만 아니라 인간의 언어정보처리 원리와 이해를 위한언어학과 뇌인지 언어정보처리 분야까지 핵심적인 요소로 작용하고 있다. 또한 NLP에는 자연어 분석, 자연어 이해, 자연어 생성 등의 기술이 사용된다. 자연어 분석은 그 정도에 따라 형태소 분석(morphological analysis), 구문 분석(syntactic analysis), 하나로써 문장의 의미에 기저(基底)하여 그 문장을 해석하는 시맨틱 분석(semantic analysis)과 문장이 실제로 무슨 의미를 내포하는지 결정하는 실용 분석(pragmatic analysis) 등으로 크게 나누어 구분할 수 있다.

2.3.2. NLU(Natural Language Understanding)

NLU는 쉽게 ‘자연어 이해’에 해당한다. 즉 컴퓨터는 기계어로만 언어를 이해할 수 있기 때문에 사람의 언어(자연어)를 기계의 언어로 이해할 수 있게 변환시키는 것을 뜻한다. NLU 기능의 예로는 문장의 의도 분류, 서로 다른 언어 간 번역 문장 생성, 자연어 질문에 대한 답변 추출 등이 있다.

딥러닝 기기반의 NLU 방식은 데이터로부터 특성(Feature)을 자동으로 학습하는 방식이다. 기존보다 폭 넓은 문맥 정보처리가 가능하고, 사진,

음성 등과 같이 다른 분야의 모델들과 연결한 멀티모델 구축이 쉬워진 장점이 있다. 또한 기존에 학습하지 않은 신조어나 오타에 robust한 처리 기술이 있어 룰 기반 방식의 한계를 보완할 있다.

STT(Speech to text)기술을 활용하여 텍스트로 만든 발화 데이터가 있거나 챗봇 대화 로그에 쌓인 데이터들을 컴퓨터가 인식하고 계산할 수 있도록 정량화 한다. 임베딩(Word embedding)기술을 활용하여 단어나 형태소를 벡터화 하여 문서를 매트릭스로 변환하는 작업을 할 수 있다. 이렇게 되면 비로소 문장 감성분류나, 기계 번역 등과 같은 업무를 딥러닝 모델로 학습할 수 있다.

이후에 문장 분류(Sentence Classification)와 같은 기술을 활용하고, 더 나아가서 한 문장 또는 문서가 입력될 때 기계가 다른 문장을 출력하는 Sequence to Sequence 기술을 활용하여 기계번역에 사용할 수 있다.

2. 4. 챗봇 보안에 관련하여

인공지능 챗봇 서비스를 더욱 이용자에 맞게 제공하고 적절한 결과를 추출하기 위해서는 많은 데이터가 필요하지만 이러한 데이터가 개인의 정보를 포함하여서는 안 된다.

IT는 시스템에 예상치 못한 데이터가 입력되는 상황에 대비해야 하고, CSO는 이 정보를 보는 사람이 누구인지도 파악해야 한다. 데이터가 필연적으로 이동하는 곳이 어디인지를 파악(로컬, 클라우드 등) 해야 하고, 챗봇은 어떻게 훈련되는지를 파악해야 한다.

직원이 검토 할 때 “데이터를 보는 사람은 누구인지, 서비스가 밖으로 유출되는지, 집단 파일에서 벗어나는지, 문제에 관심을 쏟고 있는지” 등에 대하여 모두 파악해야 한다.

챗봇 위험에 관한 SLA(Service Level Agreement)를 이행하는 것도 하나의 방법이라고 말한다. 일반적으로 SLA에 포함되는 업타임(Uptime) 요건, 품질 기대치, 기타 사안을 포함하는 것 외에 암호화 등의 보안이 필요하다.

위험을 완화하기 위해서는 챗봇 서비스 진행시 축적되는 데이터에 대하여도 검증이 필요하다. 제공하는 서비스가 개인정보를 얼마나 포함하는지, 개인의 사소한 부분까지 기록을 하게 되는지, 데이터를 얼마나 축적하게 되는지에 대한 꼼꼼한 검토가 필요하다. 이러한 데이터는 직관적일수록 좋다고들 한다.

미국에서 방문한 UMS는 개인고객 데이터를 아주 많이 다루고 있다고 한다. 구체적인 보안관리 방안을 기술할 수는 없지만 개인정보를 다루고 있는 만큼 보안을 아주 엄격하게 관리한다고 하였다. 기본적으로 2중 보안은 기본으로 적용하고 있으며, 보안업체의 선정부터 하여 이러한 데이터가 외부로 유출되지 않도록 자체적인 보안솔루션을 적용하고 있다.

3. 인공지능 챗봇서비스 활용 사례

3.1. 국외 챗봇서비스 활용 사례

3.1.1. 미국농무부(USDA)의 “음식 안전 및 검사 서비스(Ask Karen)”

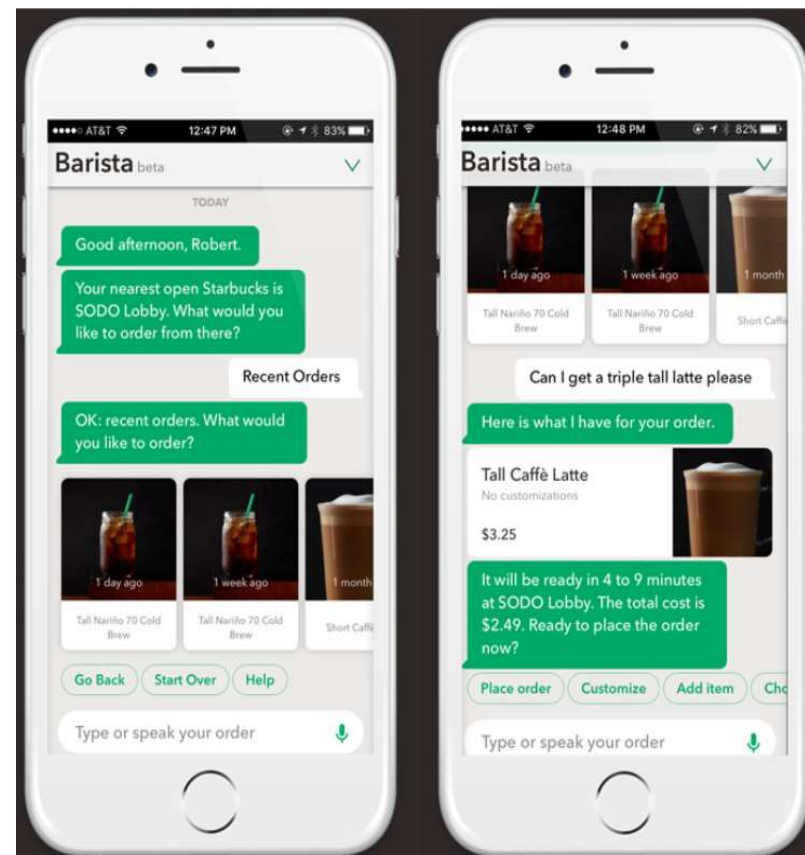
Ask Karen이라고 명명된 동 서비스는 메신저 기반의 챗봇 서비스로서 언제 어디서나 메시지를 통해 신뢰도 높은 음식 안전 및 검사 서비스 제공한다. 음식 안전과 검사 서비스를 제공받기 위한 기존의 수많은 절차들을 통합·연계하여 각 절차의 수행에 따르는 비용 절감하는 효과가 있다.

메신저라는 공통의 플랫폼을 통하여 서비스를 제공함으로써 타 부서의 서비스 제공시 이용자에게 동일한 UI/UX 경험 제공 가능하다.

* Ask Karen 서비스는 현재 확인 결과 직원연결을 통한 채팅서비스로 변경되어 제공되고 있다.

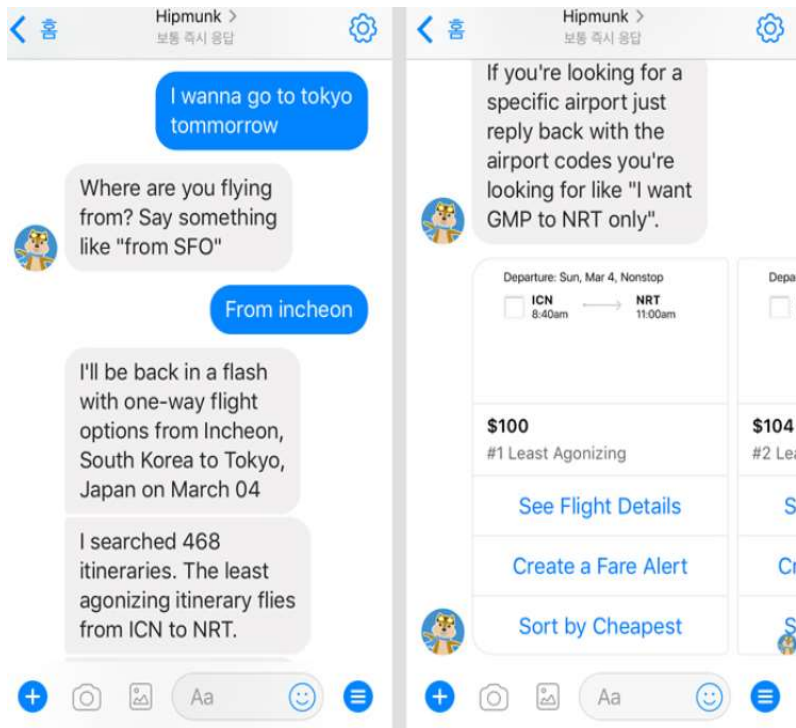
3.1.2. 마이스타벅스 바리스타

스타벅스에서 주문을 할 수 있는 챗봇이다.. 'Can I get a triple tall latte please'처럼 입력하면 문장을 이해하고 거기에 맞는 커피를 바로 예약할 수 있고, 지정한 매장에 가서 계산을 하고 받으면 되는 시스템이다.



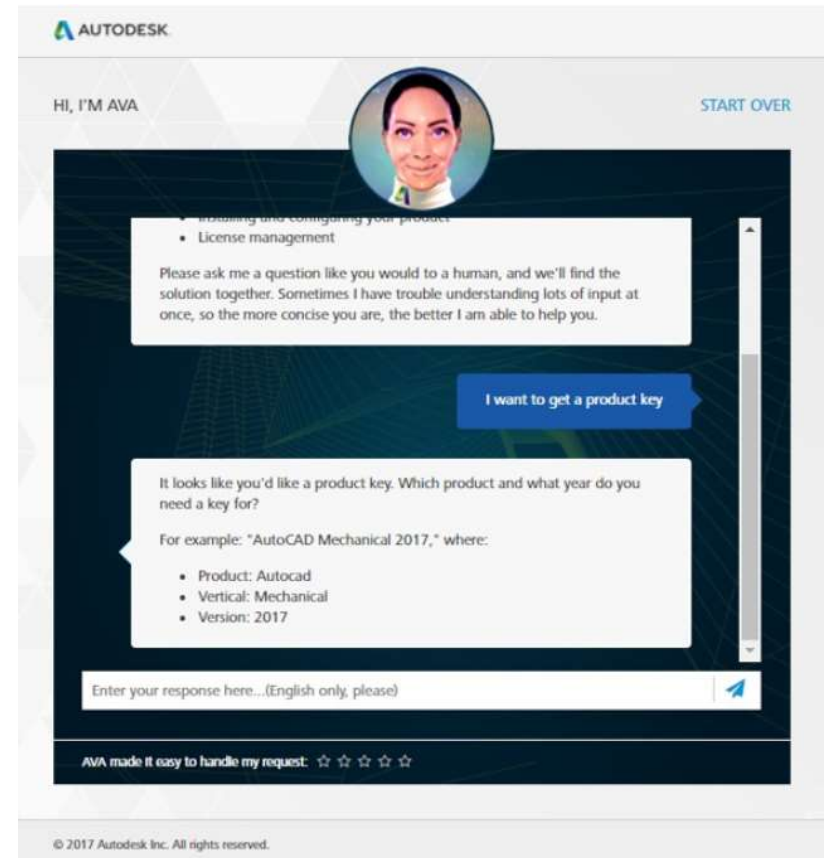
3.1.3. Hipmunk

Hipmunk는 비행기, 호텔, 렌트카 등 여행에 관련된 정보를 제공하고 예약까지 할 수 있는 챗봇으로 '도쿄로 내일 가고 싶다'고 입력하면 어디서 출발하는지 다시 물어보는 등 서로 대화를 주고받으면서 원하는 정보를 찾을 수 있다.



3.1.4. AVA(Autodesk Virtual Agent)

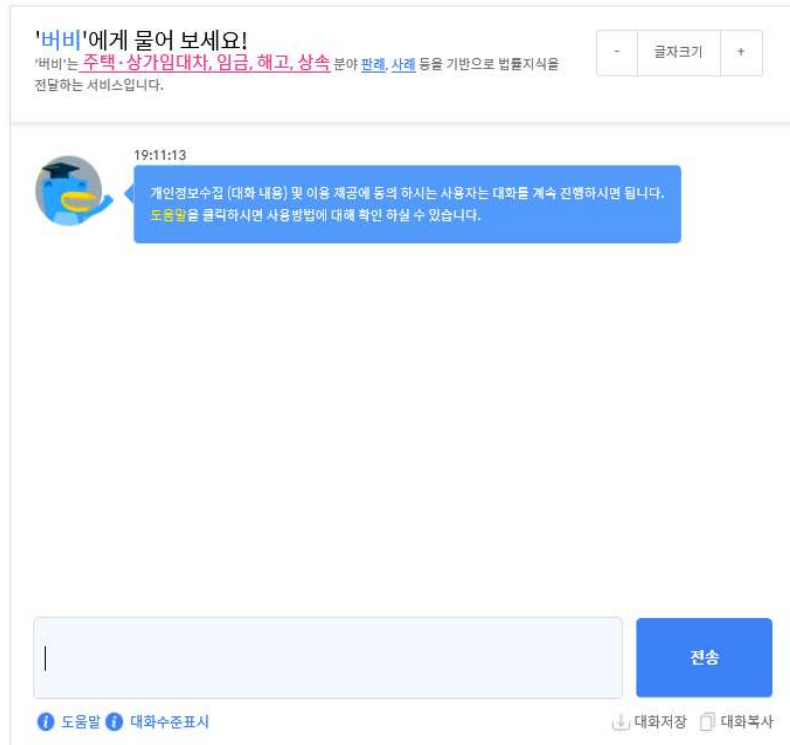
오토데스크에서 운영하는 챗봇으로 AutoCAD, 3ds Max 등 여러 제품에 대한 서비스를 제공한다. 프로젝트 키나 프로그램에 대해 궁금한 점을 대화를 통해 물어볼 수 있다. 제퍼디 퀴즈에서 사람을 상대로 우승하기도 했던 IBM의 왓슨을 사용한 시스템이다.



3.2. 국내 챗봇서비스 활용 사례

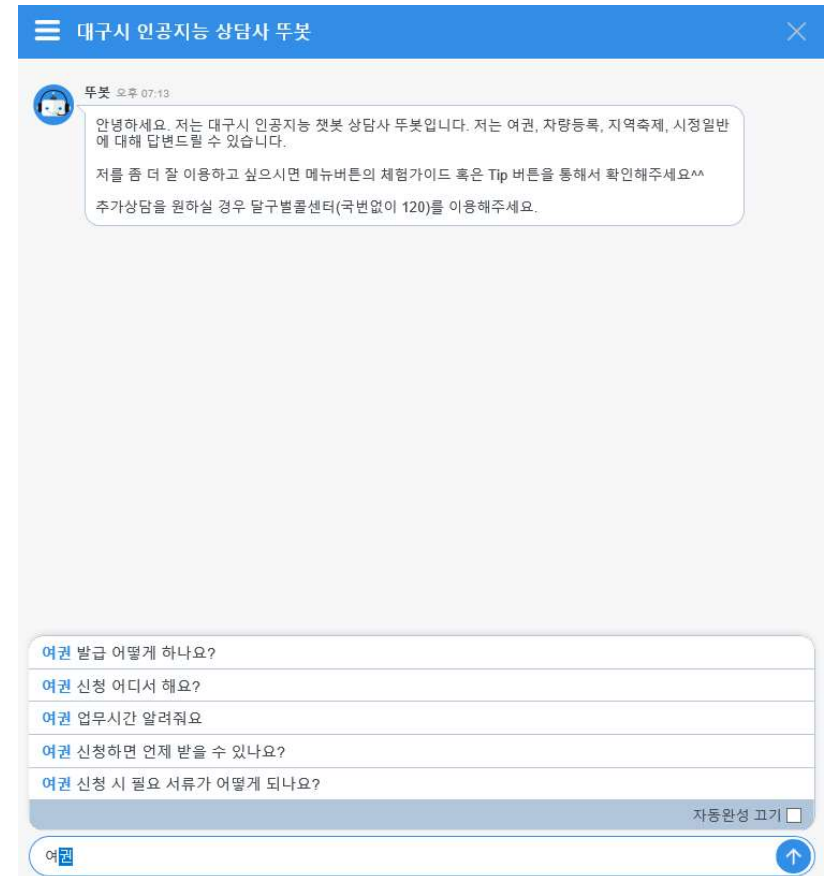
3.2.1. 법무부의 “버비”

쉽고 재미있게 생활 법률 서비스를 제공하기 위해 대중적인 메신저(카카오톡, 라인 등)를 매개로 인공지능 기반 대화형 법률지식 서비스 제공한다. 부동산 거래, 상속, 교통사고, 임금퇴직금 등 생활 법률 지능정보 서비스를 언제 어디서나 챗봇 서비스인 버비를 통해 확인 가능하다.



3.2.2. 대구시 “뚜뚱”

여권, 차량등록, 지역축제, 시정일반 등에 대한 민원사항에 대하여 언제든지 문의하고 응답받을 수 있도록 인공지능 서비스를 통하여 제공하고 있다. 단순반복 민원응대에 따른 담당자 업무과중 절감과 민원 대기시간 축소를 통해 업무 효율성 및 민원인 만족도를 제고할 수 있다.



III. 결론

인공지능은 처음 용어가 정의되고 세상에 알려진 이후로 아주 많은 발전을 하였고, 앞으로도 무한한 발전의 가능성을 가지고 있다. 현재만 해도 과거에 꿈꿔왔던 사람이 운전하지 않고 기계가 운전하는 자율주행서비스가 상용화되기를 기다리고 있는 단계이고, 로봇이 커피를 주문받고 서비스하는 시대가 왔다. 이러한 서비스들은 앞서 기술한 여러 가지 기술이 유기적으로 연결되고 접목되어 가능하다. 인공지능을 활용한 기계학습(Machine Learning) 기술, 기계가 학습을 더욱 심도 있게 할 수 있게 하는 기술인 딥러닝(Deep Learning), 자연어 처리 기술(NLP, NLU), 위에는 기술하지 않은 텍스트 마이닝(Text Mining) 등 아주 많은 인공지능 기술들이 존재하고, 어떠한 기능을 어떤 방식으로 활용하느냐에 따라 무궁무진한 활용이 가능해진다.

인공지능을 활용한 발전 중 대부분이 우리의 일상생활과 밀접하게 관련이 있다. 그 중에서 인공지능 기반의 ‘챗봇’은 몇 년 전부터 이미 제공되고 있거나 앞으로 제공하기 위하여 개발 중인 서비스이다. 스마트폰의 발전으로 언제 어디서든 인터넷을 사용할 수 있고, 대화할 수 있는 어플리케이션이 제작되어 널리 활용되고 있다. 때문에 챗봇 서비스의 발전을 통하여 우리의 생활은 더욱 편리해 질 수 있다. 상담원을 통해 문의해야 하던 내용들을 한 두 번의 채팅만으로 해결 할 수 있고, 많은 정보를 담아 복잡해진 홈페이지를 전부 둘러보지 않아도 원하는 서비스를 간단히 안내받아 이용할 수 있게 된다.

정보가 넘쳐나고 서비스가 편리해 질수록 유의해야 하는 부분이 있다. 바로 보안 분야이다. 개인에게 적합하고 맞춤형 서비스를 제공받기 위해서는 각 개인의 정보를 제공하는 것이 불가피하다. 검색포털 서비스에서 내가 몇 번 검색하였던 제품이 있을 때 그것과 유사하거나 관련 있는 제품이 포털 홈페이지 광고에 자주 나타나는 것을 경험 한 적이 다들 있을 것이다. 여기서 나도 모르게 제공되는 정보들이 있기 때문에 포털 등에서 내가 검색했던 것과 관련 있는 광고를 보여주게 되는 것이다. 이때 제공

되는 개인의 정보들의 보안이 엄격하게 관리가 되지 않고 유출된다면 정보의 호수 속에 내 정보가 포함되어 많은 부분에서 피해를 보게 될 수 있다. 따라서 인공지능 서비스를 제공하면서 연계 되는 정보 데이터들은 아주 엄격하게 관리하고 보안에 더욱 유의하여야 한다. 이는 컴퓨터 내에서 관리하고 시스템, 보안 프로그램만으로 관리가 되는 것은 아니다. 데이터를 관리하기 위한 보안정책을 알맞게 제정하는 것이 필요하고, 이를 철저히 지켜야 한다. 보안프로그램은 말할 필요도 없다. 그리고 관리자 또는 담당자가 서비스를 제공하면서 발생하는 DB를 확인하여 불필요한 정보는 정리할 필요가 있다. 이 외에 보안을 지키기 위한 많은 방안들을 고민하여 제공되는 서비스를 안심하고 이용할 수 있게 하여야 하는 것이 아주 중요하다.

기술과 보안, 연구, 관리자의 노력, 이용자의 관심 이 모든 것이 조화롭게 이루어 졌을 때 인공지능 기술은 더욱 발전하게 되고 널리 서비스 될 것이다.

□ 참 고

- 조영임, 인공지능시스템, 흥릉과학출판사, 2012
- 과학정보기술통신부(2018), I-KOREA 4.0 실현을 위한 인공지능(AI) R&O 전략
- ETR(2017), NIA(2018), IITP(2018)
- Mckinsey&Company(2018), Notes from the AI frontier insights from hundreds of use cases
- 김병희, 장병탁, 딥러닝(2017) 인공지능을 이끄는 첨단 기술
- LG CNS(2017) 부족한 데이터로 하는 머신러닝! 전이 학습
- 아이티랩(itlab.co.kr)
- NIA 한국정보화진흥원(2018) 인공지능 기반의 ‘챗봇(ChatBot)’ 서비스 등장과 발전 동향
- 김성근, 신민철, 강주영(2018), 챗봇 기술 소개 및 사례 분석
- 이창기, 김준석, 김정희, 김현기,(2014) “딥 러닝을 이용한 개체명 인식”,한국정보과학회 학술발표논문집,
- Sunnie Sun Chu(2019) Unsupervised Topic Model Based Text Network Construction for Learning Word Embeddings
- Vahid Noroozi 외 4인(2019) Enhancing Decision Tree based Interpretation of Deep Neural Networks
- Ramkumar Harikrishnakumar(2019) Supervised Machine Learning Approach for Effective Supplier Classification
- Xiang Liu외 3인(2019) Multiple Learning for Regression in Big Data
- Omid Shahmirzai(2019) Text Similarity in Vector Space Models
- Aaron R Richter 외 1인(2019) Deep Learning and Thresholding with Class-Imbalanced Big Data
- Vidhushini Srinivasan(2019) Natural Language Generation Using Reinforcement Learning with External Rewards
- 인공지능신문(www.aitimes.kr)
- 대구시 인공지능 챗봇서비스 <http://talk.lawnorder.go.kr/web/index.do>
- 법무부 인공지능 챗봇서비스 <http://talk.lawnorder.go.kr/>
- AVA(Autodest Virtual Agent) <https://ava.autodesk.com/>
- 인공지능 개발자 모임 <http://aidev.co.kr/chatbot/3768>