미래를 주도하는 대표적인 머신러닝 기업

IBM 왓슨, 구글 딥마인드, 마이크로소프트(MS), 아마존 등 머신러닝을 대표하는 세계적인 기업들의 특징에서 머신러닝의 미래를 볼 수 있다. 각 분야와 현재까지의 성과, 그리고 아키텍처와 방법론의 차이를 알아본다.

글 황문기 서강미래기술원

지난 3월 알파고와 이세돌 9단의 세기적 대결로 전 세계는 구글의 마법에 빠진 듯 인공지능AI의 세상으로 급속히 빨려 들어갔다. 지난해 구글이 텐서플로Tensorflow라는 오픈소스를 공개한 지 불과 석 달 만의 일이다. 구글은 머신러닝 분야에서 글로벌 플레이어로 부상하게 돼 이번 대국 이벤트를 통해 34조 원의 기업가치 증가 효과를 봤다고 한다. 이는 구글의 한 발 앞선 인재 영입과 오픈 이노베이션 전략이 가져다준 엄청난 성공이라고 할 수 있겠다.

머신러닝 기술 동향

마이크로소프트, 페이스북, 아마존 등 글로벌 플레이어들은 이미 2012년부터 인공기술 분야 시장을 주도하기 위해 엄청난 투자로 인재를 영입해 인공지능 플랫폼에 공을 들이고 있다. 다음은 주요 플레이어들의 머신러닝 플랫폼 추진 동향이다.

마이크로소프트 DMTK

DMTKDistributed Machine Toolkit 프레임워크는 시스템 클러스터에서 다양한 기계학습 작업을 배포하는 문제를 해결해준다. DMTK는 완전한 기성 솔루션이라기보다는 하나의 프레임워크로서 과금이 되고 실제 알고리즘의 수는 적지만 DMTK 디자인 덕분에 확장이 가능해 사용자들은 제한된 자원으로 클러스터를 활용할 수 있다.



사진설명 : 3월 15일, 독일 하노버의 정보통신 박람회 '2016 세빗(CEBIT)'에서 참가자들이 IBM사의 인공지능 로봇 '페퍼'를 살펴보고 있다.

구글 텐서플로

구글의 퀴베르메테스Kubernetes와 같이 구글 내부의 문제를 해결하기 위해 개발하여 지난해 12월 공개됐다. 텐서플로는 데이터 흐름 그래프data flow graphs라는 것을 실행하는데, 데이터 배치tensor는 일련의 알고리즘에 의해 처리된다. 시스템을 통한 데이터 이동을 플로flows라고 하는

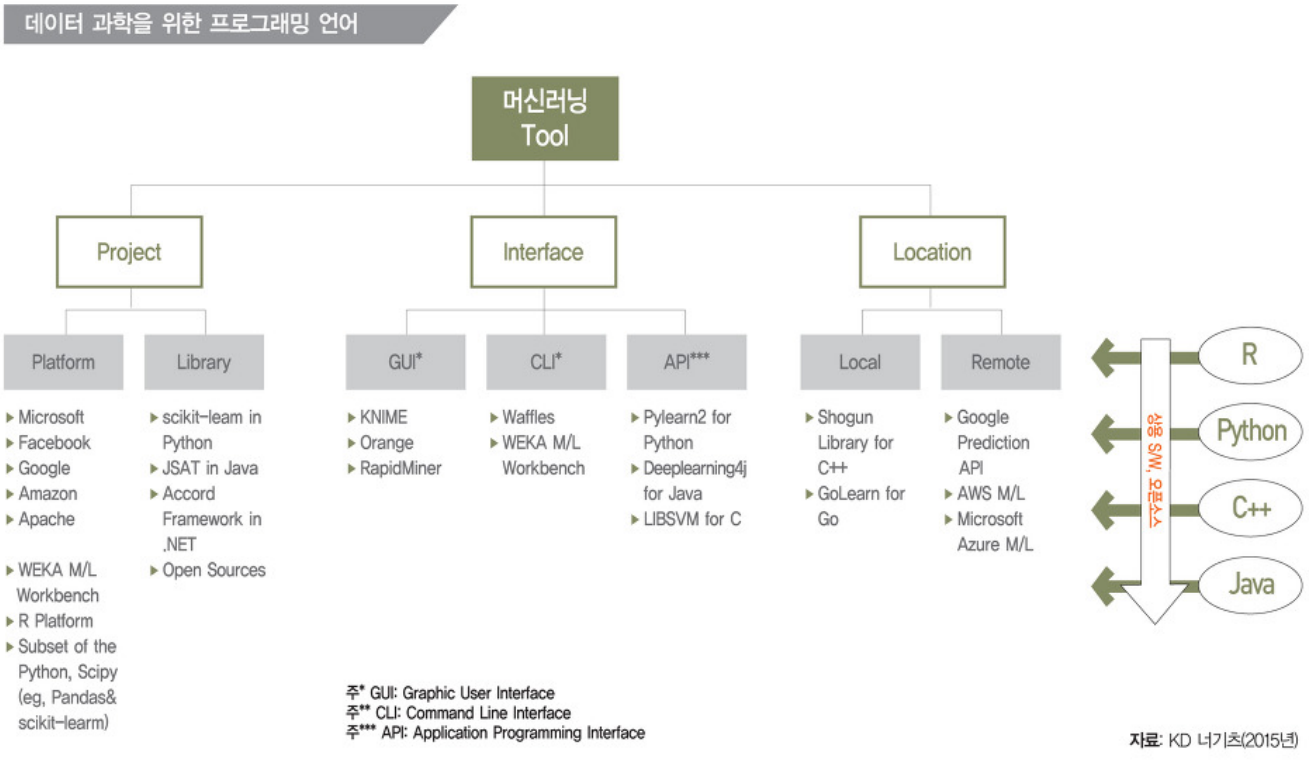
데, 텐서플로의 명칭은 여기에서 비롯됐다.

페이스북 토치

토치Torch는 오픈소스 개발 환경으로 과학 분야 컴퓨팅, 머신러닝, 컴퓨터 비전을 다룰 때 딥러닝과 뇌회로망Convolutional Neural Network, ConvNet 등 전문 분야에서 주로 사용한다. 토치는 대학연구소, 구글 딥마인드, 트위터, 인텔, 엔비디아, AMD 등 많은 기업들이 사용한다. 오픈소스 모듈은 대용량 ConvNet뿐만 아니라 자연어 처리NLP 애플리케이션에서 많이 사용하는 그래픽처리장치GPU에 최적화된 것이 강점이다.

카페

2013년 머신 비전 프로젝트를 위해 개발된 카페Caffe는 미국 버클리 캘리포니아대에서 관리하고 있으며 현재 가장 많은 사람들이 사용하고 있는 라이브러리다. 시플러스플러스C++로 직접 사용할 수도 있지만 파이선Python과 매트랩Matlab 인터페이스도 잘 구현돼 있다. 속도가 관건이었기 때문에 전체가 C++로 작성됐으며, 쿠다CUDA 가속을 지원하고 필요에 따라 중앙처리장치와 GPU 연산이 전환되기도 한다.



머신러닝 기술 구성

현재 머신러닝 기술은 대기업의 머신러닝 플랫폼과 상용SW 및 오픈소스 라이브러리 등이 공존하는 복잡한 생태계를 형성하고 있다. 미국에는 전국적인 데이터 과학data science 전문가들이 공개적으로 참여하는 프로젝트 플랫폼인 캐글Kaggle이 있다. 상용 소프트웨어와 오픈소스(MIT, 버클리 캘리포니아대, 뉴욕대 등)에서 만든 라이브러리와 상용 소프트웨어는 상호 협업하면서 개발자들의 필요와 산업별 특성에 따라 다양한 선택이 가능하다. 현재 대표적인 상용 플랫폼과 머신러닝 전문가들이 주로 사용하는 라이브러리, 인터페이스 방식, 위치 등에 따라서 정리해볼 수 있다.

증권업 머신러닝

전 세계적으로 불고 있는 인공지능에 대한 관심은 자본시장, 자산관리 분야에도 예외는 아니다. 해외에서 'AI 금융'이라 불리는 이 분야에서 자연어 처리, 감성 분석, 예측 분석과 행동재무학behavioral finance, 실시간 클러스트 등 다양한 머신러닝 기술을 활용해 전통적인 대형 금융기관 백오피스 업무에도 도전장을 내고 있다.

자연어 처리Natural Language Processing

자연어 기반의 알고리즘 기술과 딥러닝 기법을 사용해 텍스트에서 값을 추출해 뉴스와 미디어 정보, 마켓 리서치 보고서 등 금융기관 백오피스 업무와 애널리스트가 하는 역할을 수행한다.

감성 분석Sentiment Analysis

텍스트 마이닝text mining 기술로 작성자의 감정을 추출해 문서의 주제보다 감정이 어떠한지를 판단해 분석한다.

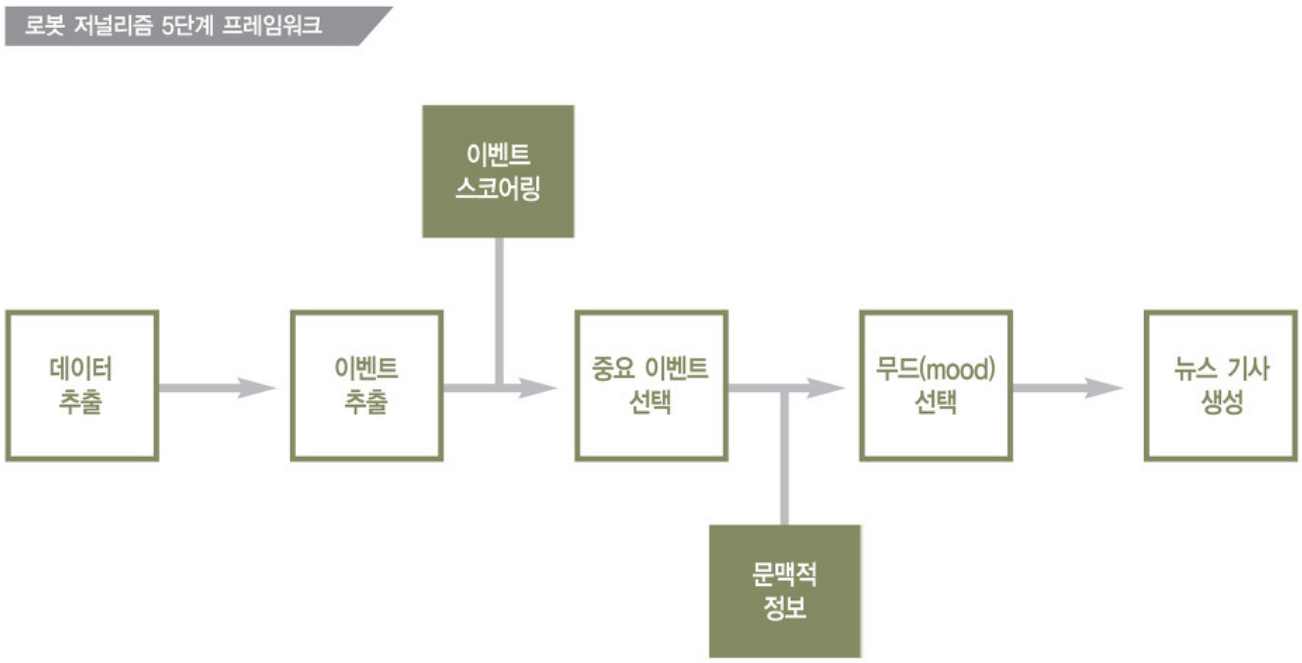
예측 분석Predictive Analytics

데이터 마이닝 기법을 통해 과거 데이터와 미래 상황에 대한 가정을 분석, 모델링해 고객 반응이나 비즈니스 활동을 예측한다.

실시간 클러스터Clusters in real-time

해외 자산관리 분야에서 머신러닝이 적용된 새로운 핀테크fintech 모델은 기존의 대형 금융기관 내

에서만 엄격하게 적용됐던 분석 기술이 이제 일반 투자자들에게도 상용화 서비스로 열리고 있음을 의미한다.



한국판 켄쇼를 기대

미국 보스턴에 위치한 켄쇼테크놀로지Kensho Technologies는 대규모 병렬 통계 컴퓨팅과 혁신적인 머신러닝 기술을 통해서 많은 기관투자가들에게 실시간 마켓 리서치 서비스를 제공하는 회사이다. 2014년 골드만삭스는 이 회사에 1500만 달러를 투자했고 CNBC는 켄쇼스테이트박스Kensho Stats Box를 서비스 중이다.

국내에도 핀테크 붐이 일면서 자본시장에도 다양한 형태의 핀테크 서비스가 소개되고 있다. 알고리즘 트레이딩, 자산운용, 투자 자문 등 다양한 로보어드바이저 서비스가 소개되고 있고 증권사들도 새로운 로보어드바이저 상품과 서비스를 출시하고 있다.

현재는 로봇 기사의 알고리즘 능력이 데이터 수집과 분석, 기초적인 내러티브 생성에 머물고 있지만 향후 자연어 처리와 머신러닝 기술이 접목되면 보다 신속하고 정확한 투자 정보 서비스도 가능할 것으로 기대한다.

미국의 켄쇼테크놀로지는 시시각각 변하는 국내외 시장 이슈와 돌발상황에 대한 실시간 마켓 리서치를 제공한다. 켄쇼테크놀로지가 시시각각 변하는 국내외 시장 이슈와 돌발 상황에 대한 분석 보고

서를 내듯이, 국내에서도 실시간 스트리밍 데이터에 머신러닝 기술이 접목된다면 증권 시황 기사뿐만 아니라 자본시장에서의 다양한 투자 정보, 기관투자자용 실시간 분석 서비스, 로보어드바이저 서비스가 가능할 것으로 예상된다. 2016년 한 해 한국에서도 머신러닝 기술을 접목한 한국판 켄쇼의 출현을 기대해본다.

로봇 기사 메커니즘의 이해를 돕는 5단계 프레임워크

데이터 추출 웹사이트나 SNS로부터 데이터를 수집하는 단계
- 웹사이트 페이지에서 불필요한 태그 제거 후 데이터 선별
- 소스 데이터 구조(data structure)에 맞춰 알고리즘 설계 및 형식 변환 필요

이벤트 추출 데이터 해석 과정
- 문화적 배경, 관찰 경험, 상황과 맥락에 따른 상황, 간접 경험, 경험적 판단
- 규칙의 알고리즘化: 예, 주가 변동, 재난 발생, 스포츠 경기 등

중요 이벤트 발견 추출된 의미 있는 이벤트에 가중치를 부여
- 흥미로운 트렌드 발견, 이상 기록 발견, 연속적 특정사건 발생 시 가중치 부여
- 알고리즘이 자동화된 상황을 판단할 수 있도록 각 이벤트에 스코어링

기사 무드 발견 중요도와 맥락을 고려해 기사의 관점을 설정
- 무드를 통해 여러 관점을 종합하고 중요한 이벤트를 엮어서 일관된 사건 전달
- 예) 경기 회복, 실업 감소 등

뉴스 기사 생성 선별된 중요 이벤트에 대한 적절한 문장 선택, 스토리 작성
- 여러 문장을 만들어 놓고 상황과 맥락에 따라 문장을 갖다 쓰는 템플릿 개념
- 알고리즘은 템플릿 속성값과 매칭 스코어를 계산해 최적의 문장을 선택

* 저작권법에 의하여 해당 콘텐츠는 코스콤 홈페이지에 저작권이 있습니다.

* 따라서, 해당 콘텐츠는 사전 동의없이 2차 가공 및 영리적인 이용을 금합니다.