

브레인 클라우드

Brain's Pick

최근 부상하고 있는 자율주행 기술부터 질병 진단, 재난 구조 로봇에 이르기까지 많은 분야에 걸쳐 인공지능이 핵심적인 역할을 담당 할 것으로 기대된다. 이에 따라 인공지능을 미래 먹거리로 인식하는 전 세계 많은 기업들이 너도나도 관련 기술 개발과 인재 유치에 사활을 걸고 있으며, 인프라 투자에도 대대적인 지원을 아끼지 않고 있다.

인공지능의 주요 분야인 딥러닝(deep learning)의 연구 과정은 대략 다음과 같다.

딥러닝의 연구 과정

- (1) 학습 알고리즘을 설계한다.
- (2) 데이터로 알고리즘을 학습시킨다.
- (3) 알고리즘이 내놓은 결과와 정답 사이 오차를 계산해 알고리즘을 수정한다.

그런데 이 연구를 진행하는 동안 연구자가 느끼는 고충이 하나 있다. 바로 학습 결과에 따라서 이 과정을 수십 번 혹은 수천 번 반복해야 한다는 점이다. 빠르게 실험할 수 있는 인프라를 갖춘다면 짧은 시간 안에 더 좋은 결과를 낼 확률을 높일 수 있을 것이다. 실험에 들이는 ‘시간’을 줄이는 것 자체가 딥러닝의 핵심 경쟁력이다.

인공지능 연구소인 카카오브레인은 연구자들이 딥러닝 연구에 집중하는 환경을 조성하고자 클라우드 플랫폼인 ‘브레인 클라우드(Brain Cloud)’를 구축했다. 브레인 클라우드는 응용 프로그램에서 요구하는 자원과 개발 환경을 제공한다. 아마존(Amazon)의 ‘아마존 웹 서비스(AWS)’, 마이크로소프트(Microsoft)의 ‘애저(Azure)’와 같은 상용 클라우드 플랫폼을 생각하면 된다.

[그림 1] 브레인 클라우드 메인 화면 ©카카오브레인



클라우드 플랫폼을 개발한 이유

프로그램 개발에서 말하는 인프라는 바로 컴퓨팅(computing) 자원이다. 우리가 일상에서 마주하는 공간에 가까이 두거나 항상 휴대하는 노트북, 데스크톱, 스마트폰이 컴퓨팅 자원의 단적인 예다. 이 글의 본문에서 가리키는 컴퓨팅 자원은 서버(server)다. 거대한 컴퓨팅 자원만을 활용하려는 목적에서 네트워크 접속 기능을 세팅해 서버실과 같은 장소에서 상시 운영하는 컴퓨터를 가리킨다. 일반적으로 클라이언트에게 정보와 서비스를 제공하려는 목적에서 서버를 둔다.

기존에는 이 컴퓨팅 자원의 사용량을 예측해서 미리 준비해놔야 했다. 문제는 예측량이 사용량보다 많으면 유휴

자원이 발생하게 되고 예측량보다 사용량이 더 많으면 서비스 운영에 문제가 생긴다는 점이다. 새로운 서버를 갖추거나 기존 서버를 재활용하기도 쉽지 않다. 프로젝트가 요구하는 환경에 맞게 서버를 변경하고 적합한 하드웨어 성능이 구비되어 있는지 일일이 확인해야 하기 때문이다. 이마저도 관리자가 프로젝트 관리자와 직접 소통해가며 각각의 서버와 프로젝트 상태에 따라 자원 할당을 조절할 수 밖에 없다. 이는 클라우드 플랫폼의 가상화(virtualization)와 오케스트레이션(orchestration)을 통해 해결할 수 있다.

카카오브레인 설립 초기에 합류한 구성원들은 자체 클라우드 플랫폼의 중요성을 절실히 깨닫고 두가지 이유에서 외부 클라우드 플랫폼을 이용하는 대신 직접 개발하기로 결정했다. 첫 번째 이유는 상대적으로 적은 운영 비용이다. 아마존 웹 서비스와 같은 외부 클라우드 플랫폼을 사용하면 자체 플랫폼 대비 5배 이상의 비용이 든다. 아울러 외부 협업에 사용되는 막대한 양의 컴퓨팅 자원을 감당하기 위해서는 직접 클라우드 플랫폼을 운영하는 편이 비용 부담이 적었다. 또 다른 이유는 딥러닝 연구에 최적화된 기능만 사용하기 위함이었다. 외부 플랫폼은 일반적인 서비스 개발에 초점을 둔 기능을 제공한다. 반면, 자체 개발한 플랫폼은 딥러닝 연구에 필요한 기능을 집중적으로 지원할 수 있어 보다 효율적인 연구 환경을 조성하는 데 큰 도움이 된다.

카카오브레인은 여러 시행착오를 겪으며 클라우드 플랫폼 개발에 착수했다. 하지만 플랫폼 개발 경험도 부족하였고, 각자 맡은 역할이 따로 있던 상황이라 클라우드 플랫폼 개발에 난항을 겪었다. 이런 와중에 카카오에서 컨테이너 기반 클라우드 플랫폼인 DKOS(Data center of Kakao OS) 개발을 주도한 최규동 개발자가 카카오브레인에 합류했다. 그의 영입 이후, 클라우드 플랫폼 개발은 큰 진척을 보이기 시작했다. 최규동 개발자를 주축으로 한 클라우드 개발팀은 우선 그래픽 처리 장치(graphics processing unit, GPU) 장비의 사용률을 극대화하는 데 집중했다. 아울러 연구자의 클라우드 사용 편의성과 연구 속도를 높여주는 다양한 응용 프로그램 지원에도 총력을 기울였다.

DKOS(Data center of Kakao OS)

지금은 많은 서비스가 컨테이너로 운영되고 있지만 그 당시만 하더라도 국내에서는 굉장히 새로운 시도로 인식되었기 때문에 일부 신규 서비스에만 DKOS를 적용했다. 이후 많은 개선을 통해 높은 안정성을 확보했고 현재는 다음 메일을 비롯한 많은 카카오 서비스들이 DKOS를 통해 제공되고 있다.

글 | 이수경 samantha.lee@kakaobrain.com

2016년 3월 알파고와 이세돌 9단이 펼치는 세기의 대결을 두 눈으로 목도한 이후 인공지능을 제대로 공부해 봐야겠다고 결심했습니다. 인공지능 본진이자 연구소인 카카오브레인으로 걸어들어온 이유입니다. 인공지능 기술과 이로 인해 바뀔 미래 사회를 다루는 글을 통해 사람들과 소통하고 싶습니다.

글/감수 | 최일지 june.one@kakaobrain.com

언제나 흥미로운 것을 찾아다니는 호기심 많은 개발자입니다. 수년 전 머신러닝을 접한 이후로 관련 엔지니어링 스킬을 익히고 있으며 현재는 카카오브레인에서 인공지능을 개발하고 있습니다. 앞으로 인공지능을 활용한 상상의 현실 만들기에 힘을 보태고 싶습니다.

글/감수 | 최규동 dave.go@kakaobrain.com

시스템을 효율적으로 개선시키는데 관심이 많은 개발자입니다. 현재 카카오브레인에서는 인공지능 클라우드 플랫폼 프로젝트를 진행하고 있습니다. 인공지능 개발자들이 연구에 집중할 수 있는 환경 조성에 앞장서고자 합니다.

글/감수 | 백운혁 clint.b@kakaobrain.com

비전공 분야에 더 많은 사랑과 애정을 쏟고 있는 야행성 개발자입니다. 사랑스러운 딸의 성장을 통해 덩달아 인문학적인 감성을 품은 공학자 아빠로 쑥쑥 자라는 중입니다. 카카오브레인에서는 공학과 인문학의 균형있는 배움과 경험을 토대로 인공지능을 연구합니다.

브레인 클라우드의 기능

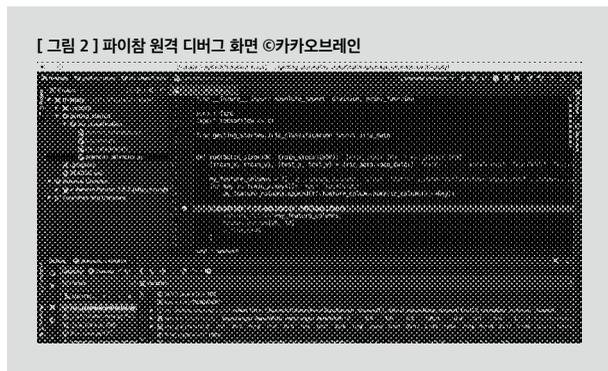
브레인 클라우드는 클라우드 플랫폼에서 제공하는 기본적인 기능뿐 아니라 연구에 특화된 다양한 기능도 제공한다. 이런 브레인 클라우드의 특 장점은 다음과 같이 크게 세 가지로 꼽을 수 있다.

첫 번째, 브레인 클라우드는 SSH(secure shell, 시큐어 셸)¹⁾를 통해 개발자가 필요로 하는 자원을 제공한다. 이를 통해 원격 IDE(integrated development environment, 통합개발환경)²⁾을 지원함으로써 연구자들이 익숙한 개발 환경에서 연구할 수 있게 해준다. 두 번째, 딥러닝에 주로 활용되는 다양한 연구 환경을 제공한다. 이는 연구자가 본인의 연구 환경을 반복적으로 사용하거나 내·외부 협업자와 공유할 때 연구 환경을 세팅하는 시간을 월등하게 줄여준다. 세 번째, 다양한 실험을 위해 여러 개의 프로그램을 병렬적으로 동시에 실행하거나 하나의 프로그램을 구성하는 부분을 분할해 동시에 처리하는 기능을 제공한다. 이는 다양한 실험 결과를 빠르게 확인할 수 있도록 해준다.

브레인 클라우드는 대용량의 고속 네트워크 파일 스토리지(file storage)로 실험 데이터를 관리하거나 사용하기에도 용이하다. 또한, 실험 결과나 과정을 메신저를 통해 알려주거나 학습에 필요한 데이터나 학습 결과 모델을 관리해주는 등 연구를 돕기 위한 다양한 기능들이 있다.

인스턴스 서버(instance server)

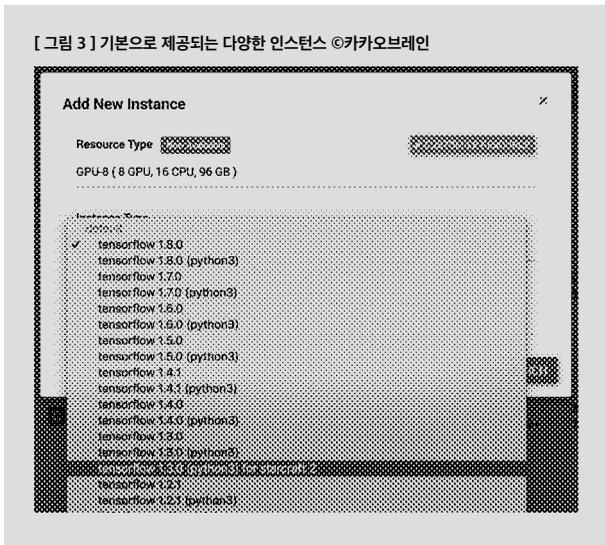
인스턴스 서버는 2개의 CPU(central processing unit, 중앙처리장치), 128GB 메모리와 같이 사용자가 원하는 만큼의 자원을 할당한다. 하드웨어에 종속되지 않기에 하드웨어적인 장애가 발생하더라도 안정적인 개발 환경을 지속해서 제공해준다. 아울러 파이참(Pycharm)³⁾이나 비주얼 스튜디오 코드(visual studio code)와 같은 IDE의 원격 개발 환경을 구축하고, 기존의 방법이나 도구를 활용한 개발 활동을 지원한다.



[그림 2] 파이참 원격 디버그 화면 ©카카오브레인

인스턴스 유형(instance type)

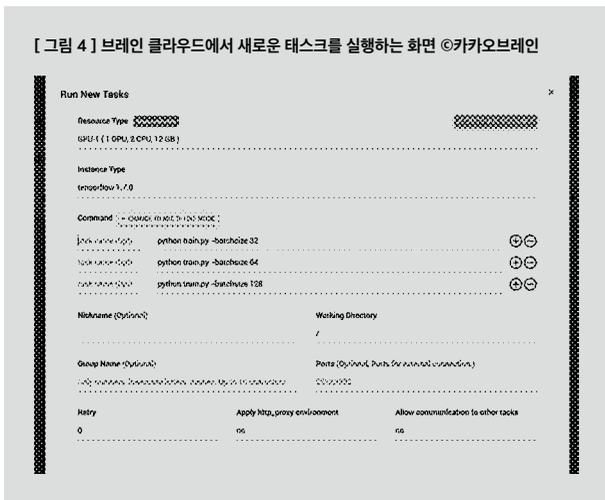
브레인 클라우드는 카페(Caffe)⁴⁾, 테아노(Theano)⁵⁾, 파이토치(Pytorch)⁵⁾, 텐서플로(Tensorflow)⁶⁾ 등 딥러닝 연구에 필요한 개발 환경을 여러 버전으로 제공한다. 연구자들은 자신이 원하는 연구 환경을 즉시 설정할 수 있다. 또는 본인이 원하는 패키지를 설치하거나 연구에 필요한 데이터 셋(data set)을 구축할 수도 있다. 이렇게 구축된 개발 환경은 반복적으로 사용하거나 다른 연구자와도 공유할 수 있다.



[그림 3] 기본으로 제공되는 다양한 인스턴스 ©카카오브레인

스케줄드 태스크(scheduled task)

스케줄드 태스크는 브레인 클라우드의 유저자원을 효율적으로 활용해 다양한 실험을 동시에 실행하고 그 결과를 빠르게 확인할 수 있게 해준다. 여러 개의 프로그램을 병렬적으로 실행하거나 하나의 프로그램을 구성하는 여러 기능을 분할해 동시에 처리하는 기능이 대표적인 예다.



[그림 4] 브레인 클라우드에서 새로운 태스크를 실행하는 화면 ©카카오브레인

앞으로도 분산학습 고도화, 최적의 하이퍼 파라미터(hyper parameter)⁷⁾ 자동 검출 등과 같이 AI(artificial intelligence, 인공지능) 연구와 관련된 더 많은 기능을 개발할 계획이다. 분산 트레이닝은 실험 속도를 월등히 높여 단시간에 실험을 끝낼 수 있게 해주고, 최적의 하이퍼 파라미터 자동 검출은 실험 결과를 자동으로 분석함으로써 단시간 내에 더욱 좋은 결과를 얻을 수 있도록 도와준다.

브레인 클라우드의 가치

지난 시리포트 11호에서 소개했듯이 카카오브레인의 연구는 자율, 공개, 협력을 원칙으로 한다.⁸⁾ ‘자율’은 연구원들이 자유롭게 주제를 선정하되 스스로가 세운 원리 원칙을 토대로 연구함을 의미한다. ‘공개’는 연구 과정과 성과를 카카오브레인 내외로 공유함을 뜻한다. 마지막으로 ‘협력’은 카카오뿐만 아니라 관련 학계 및 업계와의 협업을 극대화함으로써 대한민국 인공지능 생태계에 기여함을 가리킨다. 이를 아우르는 핵심이 바로 개방형 연구(open research)다. 브레인 클라우드는 바로 이 개방형 연구를 위한 초석인 셈이다.

일반적인 학교 연구실의 딥러닝 연구 환경은 열악한 축에 속한다. 컴퓨팅 성능이 떨어지고 이마저도 연구실에 놓인 컴퓨터 한 대를 연구원들이 돌아가며 사용하는 경우가 허다하다. 새로운 주제로 연구를 진행할 때마다 개발 환경을 세팅하는 데 많은 시간을 들여야 하는 일도 부지기수다. 학교 연구실이 처한 연구 환경을 조금이라도 개선하는데 보탬이 되고자 카카오브레인은 지난해 4월 CSI(Center for Super Intelligence, 초지능연구센터)와 산학협력을 맺고, 센터에 소속된 연구실에 브레인 클라우드를 오픈했다. 외부 연구실에 그래픽 처리 장치 자원과 다양한 기능을 제공함으로써 협업 속도를 끌어올리는 효과 또한 거둘 수 있었다.

보안 등의 이슈로 카카오브레인 클라우드를 외부에 완전히 개방하기는 현실적으로 어려운 일이었다. 이에 카카오브레인은 대학 연구실이나 기관, 기업과의 연구 협업을 위해 브레인 클라우드를 외부와 제한적으로 공유하고 있다. 이를 위해 기술적으로 해결해야 할 이슈들이 몇 가지 있었다. VPN(virtual private network)과 계정을 발급하는 방식으로 자격을 갖춘 이만 브레인 클라우드를 사용할 수 있도록 했으며, 데이터 보안을 유지하고자 계정마다 서로 다른 접근 권한을 제공하고 있다. 이를테면 B라는 연구실에서 A 연구실 데이터에 권한 없이 액세스할 수 없다.

클라우드가 곧 여러 AI 서비스를 잇는 또 다른 플랫폼인 만큼 개발팀은 연구 결과를 API(application programming interface, 응용 프로그램 프로그래밍 인터페이스)나 SDK(software

development kit, 소프트웨어 개발 키트) 형태로 공유하는 플랫폼 환경 구축을 최종 목표로 하고 있다. 클라우드 컴퓨팅 자원으로 연구하는 ‘트레이닝 하우스(training house)’와 연구 결과를 공유하는 ‘모델 하우스(model house)’ 모두를 포괄하는 클라우드 개발이 핵심이다. 트레이닝 하우스는 네이버(Naver) 클라우드 플랫폼, 아마존 웹 서비스, 구글(Google) 클라우드처럼 연구를 위한 개발 환경을 카카오브레인 얼라이언스(alliance)에 제공하는 데 의의가 있다. 모델 하우스는 네이버 클로바(Clova)와 구글 클라우드 API처럼 누구나 연구 결과를 재활용할 수 있도록 한다. 공동 연구를 할 수 있는 공간, 연구 결과를 공개하는 공간, 그리고 연구 결과를 공유하는 공간이 바로 브레인 클라우드가 제공하려는 핵심 가치라고 보면 된다.

물론 자체적으로 클라우드 플랫폼을 구축하고 이를 직접 운영, 개선해나가는 일이 쉽지만은 않다. 그럼에도 불구하고, 복잡하고 귀찮은 부분을 간소화해 더 중요한 일에 개발자가 집중할 수 있는 환경을 조성하는 것은 매우 가치있는 일이라 할 수 있다. 단순히 컴퓨팅 자원을 빠르게 제공하는 데 그치지 않고 카카오브레인과 협업하는 연구소나 기업 간 각종 노하우를 공유함으로써 모든 연구원이 보다 빠르고 편하게 연구할 수 있는 토대를 제공할 수 있어서다. 축구에 비유하자면 플랫폼은 화려하진 않지만 경기에 없어서 안 되는 ‘수비수’와 같다고 볼 수 있다. 공격수가 골을 넣는 것에만 집중할 수 있도록 수비를 튼튼하게 해주는 역할을 묵묵히 수행하기 때문이다.

¹⁾ 설명 | 원격 컴퓨터를 안전하게 제어하는 프로토콜 또는 이 프로토콜을 사용하는 프로그램을 가리킨다. SSH 클라이언트와 SSH 서버 사이에는 강력한 암호화가 적용돼 있어 중간에 데이터를 가로채더라도, 해석이 어려운 암호화된 문자만 노출된다. ²⁾ 설명 | 코딩, 디버깅, 컴파일, 배포 등 프로그램을 개발과 관련된 모든 작업을 하나의 프로그램 안에서 처리하는 환경을 제공하는 소프트웨어다. ³⁾ 참고 | 파이선(Python)이라는 언어로 소프트웨어를 개발하는 도구 중 하나다. ⁴⁾ 참고 | C++ 기반 딥러닝 프레임워크다. ⁵⁾ 참고 | 파이선(Python) 기반 딥러닝 라이브러리이다. ⁶⁾ 참고 | 구글이 만든 딥러닝 라이브러리. 여러가지 언어를 지원하며 그 중 파이선이 메인이다. ⁷⁾ 참고 | 개발자가 직접 설정해줘야 하는 값이다. 각 층의 뉴런 수, 배치 크기, 매개변수 갯수 등의 학습을 과 가치치 감소를 등이 여기에 해당한다. ⁸⁾ 참고 | 카카오 시리포트 11호 중 <카카오브레인의 1년, 그리고 미래> <https://brunch.co.kr/@kakao-it/223>