AI 이노베이션 레포트

21600685 조예성

자사의 서비스에 가장 알맞은, 학습 데이터를 선정하고 그에 맞는 알맞은 알고리즘을 선정하는 법은 매우 중요하다. 우리가 이러한 서비스를 제공하려고 하는데 그러기 위해서는 어떤 데이터가 필요한가? 이 질문이 가장 중요한 질문이라 할 수 있고 그러한 질문에 이어 그렇다면 이러한 학습 데이터가 주어지면 어떠한 모델링, 학습 기법을 사용하여 AI 모델을 구축할 것인가? 이러한 알고리즘의 선정 방법이 그 다음의 고민이라 할 수 있다. 데이터 마이닝을 하여 모델링을 구축할 때 실제적으로 가장 시간이 많이 드는 것도 데이터를 확보하고 전처리하는 것이다. 데이터를 가져올 수 있는 다양한 소스들을 파악하고 유용하게 사용할 수 있는 잠재된 raw data를 어떻게 사용할 것인가에 대한 인사이트를 갖는 것이 중요한 문제라고 할 수 있다. 이러한 데이터를 끊임없이 관리해주어 노후화됨을 방지해야 하고 dirty 데이터들을 시시각각으로 정제하여 본인의 서비스에 맞는 알맞은 데이터들 지속적으로 공급해주어야 한다. 하지만 교수님께서 강의에 언급하셨듯 이러한 데이터는 기본적으로 방대한 양을 갖고 있어야 하는데, 큰 기반이 갖추어지지 않은 기업들은 이러한 데이터의 확보가 어려울 수 있는데 이를 제공해주는 다양한 서비스들의 사용 (영상 이미지 자동 태킹 서비스 – scale api, AI 어플 기능들 자동 시행 활동 데이터 생성 플랫폼 – apptest.api) 하거나 자사에 있는 아직, 잠재성이 뛰어나나 발견하지 못한 데이터들을 유심히 살펴볼 필요가 있다. 데이터가 없는 기업은 없으므로 본인들의 데이터를 충분히 잘 확보하여 이를 잘 가공하면, 양이 엄청나게 많지 않아도 질 좋은 데이터를 확보하여 보다 나은 양질의 서비스를 구축할 수 있게 된다. 그리고 이렇게 가공된 데이터를 잘 적용시킬 수 있는 적절한 알고리즘을 선정하여 적절한 AI 서비스를 창출하는 것이 중요한 과제가 될 수 있다.

실제로도 알고리즘들은 이미 시중에 좋은 것이 많이 공시되어 있고 다른 분야에서도 연구중인 부분이 많다. 물론 이러한 알고리즘에 대해 깊이 공부하고 연구를 해야겠지만, 만일 그러한 분야가 아니라, 이러한 알고리즘을 써먹는 기업의 입장에서는 결국, 데이터가 그 승패를 좌우한다. 아무리 좋은 식기류가 있어도 재료가 좋지 않으면 음식은 맛이 있을 수가 없다. 좋은 재료, 잘 다듬고 유심히 선정한 재료들을 알맞은 식기류와 레시피로 양질의 음식을 창조하는 것이 가장 바람직한 요리의 과정이다. 적절한 데이터의 산출과 정제는 현재 AI의 트랜드로 가는 모든 기업에 가장 중요한 과제라고 할 수 있다.