

머신러닝, 유학, 그리고 무언가를 배운다는 것



서범석

2016년 봄 유학을 앞두고 있던 그때 알파고대 이세돌의 경기가 있었다. 바둑은 잘 모르지만 언론이 연신 알파고, 머신러닝, 인공지능에 감탄하는 것을 보면서 나는 흐뭇한 기분이 들었다. 한참을 공부하던 경제학을 그만두고 통계를 더 해야겠다고 생각한 것도 순전히 머신러닝때문이었다.

머신러닝은 데이터를 통해서 가장 적합한 답을 찾아내는 통계적 방법론이다. 답이 명확하지 않을때 머신러닝 방법이 유용하다. 우리가 소위 감이라고 얘기하는 것들, 경험적으로 알고 있지만 설명하기 곤란한 것들을 머신러닝을 통해 연구한다. 가령 우리는 여러 사진을 두고 같은 사람을 쉽게 구별해 내지만, 그게 왜 같은 사람이냐고 묻는다면 별로 할말이 없다. 사진이 항상 똑같지 않기 때문이다. 어떤 때는 세상 제일 못생겨 보이기도 하고 어떤 때는 연예인해도 되겠다 싶은 게 사진이니까. 머신러닝은 여러 사진에서 패턴을 찾아내 같은 사람인지 아닌지 혹은 여러사람 중 누구에 해당하는지를 찾는 데 쓰일 수 있다.

사실 내 연구가 사진에서 사람 찾는 일은 아니지만, 어쨌든 그렇게 유학나와 머신러닝을 공부한지가 4년이 다 되어간다. 재밌게도 머신러닝을 공부하면서 종종 배움이란 것에 대해 생각을 하게 되는데, 사람이 무언가를 배워가는 과정도 머신러닝과 크게 다르지 않기 때문이다. 특히 유학하면서 새로 경험한 게 많고 삶의 태도 또한 새로이 배운 게 많다보니 더 그런 생각을 할지도 모르겠다. 마치 새로운 데이터로 모델을 훈련시키면 새로운 답이 나오는 것처럼 말이다.

미국에 와서 신선했던게 많다. 그중 하나가 여기 친구들이 학교 복도며 교실 바닥이며 아무 데나 앉고 심지어 배를 깔고 엎어져 있는 걸 보는 거였다. 익히 들었던 얘기라 처음엔 그냥 그런가 보다 했다. 그런데 막상 교실 밖에 한번 누워있어 볼까 하면 꺼림칙한 기분이 들어 하지 못했다. 삼십여 년을 그래본 적 없으니 당연한 기분 아닐까. 머신러닝으로 문제를 풀다 보면 종종 로컬최적값이라고 하는 결과에 도달한다. 이 로컬최적값은 우리가 얻고자 하는 가장 적합한 답이 아니다. 초기값이 최적의 답과 아주 다르거나 적은 양의 데이터로 모델을 훈련시키다 보면 이런 결과가 나타난다.

삼십년 동안 한국에서 훈련된 내 뇌가 옳다고 믿었던 것이 로컬최적값은 아니었는지 생각하게 된다. 조심스럽게 추측해 보건데 어릴 때부터 흙은 더럽다고 배워온 것 같다. 그런데 여기 친구들을 자주 보다 보니 교실 복도에 있을 법한 작은 흙들이 그렇게 더럽진 않을 수도 있겠단 생각이 든다.

한가지를 더 얘기해보면, 내가 있는 펜실베니아 주립대학은 동서남북 반경 20분 거리 밖이 그저 드넓은 들판뿐이다. 사람들은 이곳을 설명하기 위해 종종 “middle of nowhere” 라는 표현을 쓰기도 한다. 도대체 이 곳 사람들은 무슨 재미로 사나 생각하기도 하는데, 여기 근처에 전기를 쓰지 않고 마차를 타며 현대기술을 거부한 채 사는 사람들이 있다. 복장도 링컨하면 떠오르는 수북한 수염에 흑백 계통의 베스트 딸린 정장을 입고 다닌다. 아미쉬라고 불리는 사람들이다. 주위들은 애긴데, 이들은 스무살이 되면 일년동안 핸드폰을 받고 마을을 나가서 살아본다고 한다. 그리곤 돌아와 평생 아미쉬로 살지 아니면 마을 밖 평범(?)한 삶을 살지를 결정한다고 한다. 핸드폰 없이는 화장실 가는 것마저 심심하다고 느끼는 나에겐 상상조차 힘든 일이다.

머신러닝으로 해답을 찾고자 할때 우리는 데이터를 가지고 모형을 훈련시킨다. 이렇게 훈련된 모형은 기존 데이터 범위를 벗어난 문제에 대해 종종 이상한 답을 내놓는다. 새로운 데이터가 기존 데이터와 현저히 다르다면 그 모형은 새로운 데이터를 분석하기에는 당연히 적합하지 않다. 처음 아미쉬를 봤을 땐 그 삶이 얼마나 답답할까 하는 생각이 들었다. 하지만 평생을 도시에서 살아온 내가 그들 삶을 어찌 알 수 있을까 싶다. 그들에겐 자정이 넘도록 아등바등 논문 쓰는 내 모습이 더 답답해 보이지 않을까 생각해 본다.

지난 4년동안 유학하며 배운것들을 나열하자면 끝이 없을 것이다. 거울에 반바지를 입고 다녀도 이상하지 않구나 하는 것과, 오레오 튀김도 먹을만 하다는 것, 그리고 편도 다섯시간은 운전할 만한 거리라는 것 등등. 그리고 한편으론 아직 배워야 할 게 많다는 것도 느낀다. 4년이나 됐지만 이곳 친구들과 대화하다 보면 여전히 어색해지는 순간이 찾아오기 때문이다. 서로 표정도 태도도 유머감각도 다른 이 친구들을 완전히 이해할 수 있을지는 자신이 없다. 나도 살아온 시간이 있으니. 다만 머신러닝을 공부하다 보면, 작은 양의 데이터를 추가하는 것으론 기존 모형을 바꾸는게 쉽지 않다는 것을 알게 된다. 연구를 하면서 항상 데이터를 모으려 노력하듯, 살아가면서도 선불리 답을 내리기 보다는 데이터를 조금 더 모아보려는 노력을 아끼지 말아야겠다. 무언가를 배운다는 것은 경험에서 패턴을 찾고 패턴을 바탕으로 신념을 만들어 가는 과정일 것이다. 머신러닝이 인간의 뇌를 모방했다고 하는데, 머신러닝을 공부하면서 인간의 뇌가 어떻게 배워가는지 또 잘못 배울 수 있는지를 생각해 본다.
