

부스트캠프 추천시스템 Level 1 프로젝트 개인 회고

T4117 신희재

나는 내 학습목표를 달성하기 위해 무엇을 어떻게 했는가?

프로젝트를 시작하면서 팀회의를 통해 프로젝트의 목적을 level 1 에서 배운 딥러닝과 추천시스템 이론을 복습하고 이론을 실제로 어떻게 구현해야 하는지를 코드 레벨에서 자세하게 뜯어보는 데에 두었습니다. 그리고 저는 딥러닝 프로젝트를 한 번도 경험해보지 못했기 때문에 개인적인 목표로는 딥러닝 프로젝트가 어떤 스텝으로 이루어져있고 스텝별로 달성해야 하는 주요 목표가 무엇인지를 최대한 이해해보는 데 두었습니다.

2 주의 프로젝트 기간 중 첫 한 주를 추천시스템에서 사용되는 모델을 수식부터 차근히 이해하고 주어진 베이스라인에서 이를 어떻게 구현하는지 매칭시켜보는 것으로 보냈습니다. 그리고 이론이 코드로 구현되는 부분에서 정확히 이해하지 못한 부분을 피어세션에 공유해서 함께 논의하면서 해결해갔습니다. 예를 들자면 프로젝트 초반에 FFM 모델을 이해하는 데 어려움을 겪었는데 이는 FFM 에서 "필드"가 정확히 무엇을 가리키는지 처음에 완전하게 이해하지 못했기 때문이었습니다. 이 궁금증을 피어세션에서 공유하고 할당된 시간을 넘는 토론 끝에 FM 의 수식의 f 는 차원을 가리키는 반면 FFM 에서는 원-핫-인코딩의 칼럼을 묶는 상위카테고리를 가리키며 이 둘을 혼동했기 때문에 잘못 이해했다는 것을 알 수 있었습니다.

나는 어떤 방식으로 모델을 개선했는가?

크게 데이터 전처리, 하이퍼 파라미터 튜닝, 모델 앙상블의 세 과정으로 모델을 개선하고자 시도했습니다.

데이터 전처리에서는 비교적 효율적으로 채울 수 있는 결측치들을 채워 넣고, 문자열 자료형을 범주형 라벨로 바꿔주거나 혹은 맵핑 딕셔너리를 사용해서 보다 단순한 라벨로 바꿔주었습니다. 이 과정에서 유저별 평점이 해당 유저가 주는 평가의 성향을, 책별 평점이 책이 받는 평가의 성향을 보여줄 수도 있다고 생각해서 유저별 평점과 책별 평점 피쳐를 가공해서 추가하는 것도 시도했습니다.

하이퍼 파라미터 튜닝에서는 Learning Rate 를 우선적으로 바꿔보고, 경험적으로 가장 잘나오는 0.01 을 기본 Learning Rate 으로 삼고 Batch Size 를 키워보거나 Weight Decay 를 높여보는 방식으로 디폴트 조합보다 훨씬 잘 나오는 하이퍼 파라미터를 탐색했습니다. 하지만 뒤돌아보니 WANDB 와 같은 도구를 쓰지 않고 순전히 감에 의존해 튜닝을 했던 것은 이번 프로젝트에서 아쉬운 점이었습니다.

마지막으로 모델 앙상블은 RMSE 가 잘 나온 개별 모델에 가중치를 더 주는 방식으로 조원들의 모델과 여러가지 무작위로 섞어보았습니다. 그 결과 리더보드에서 최종 4 위에 도달할 수 있었습니다. 하지만 이 또한 특정한 가설을 가지고 앙상블을 한 것이 아니라서 프로젝트가 끝나고 보다 체계적인 앙상블 실험을 해야 한다고 회고했습니다.

내가 한 행동의 결과로 어떤 지점을 달성하고, 어떠한 깨달음을 얻었는가?

위에서 말한 여러 개선 노력 중에서 평균 피쳐 엔지니어링이 소기의 성과를 얻었습니다. 처음에는 학습 세트에서 주어진 자료에 근거해서 유저와 책의 평균 평점을 구하고 이를 유저/책 테이블에 새로운 피쳐로 추가해보았습니다. 하지만 데이터 가공에 이미 포함된 검증 세트에 대해서는 RMSE 가 1.4 이하까지

나왔지만 테스트 세트에서는 2.5 이상으로 치솟는 것에서 이러한 피쳐 가공법이 은 학습 세트에 과적합되었음을 확인할 수 있었습니다.

이어서 시도한 것은 지금까지 팀에서 가장 좋은 점수를 획득한 예측의 평가 자료를 합산하여 학습 세트에 1, 예측된 테스트 세트에 4 의 가중치를 주어 가중평균한 평균 평점 데이터를 마찬가지로 유저/책 테이블에 피쳐로 추가하는 방법이었습니다. 그리고 투표수가 매우 적은 유저/책의 평균 평점은 유의미하지 않다고 생각하여 최소한 n 번 이상 투표하거나 투표 받은 책을 대상으로만 가중 평균을 내었습니다. 그 결과 검증 세트에서 2.15 정도의 성과를 냈고 리더보드 점수에서도 유사한 점수를 얻었습니다. 하지만 원래의 가장 좋은 예측보다는 좋지 않았는데 생각해보니 원래의 예측을 정답으로 가정하고 가공하는 방식이기 때문에 애시당초 원래의 예측보다 더 잘 나오기는 어렵다는 생각이 들었습니다.

이 과정을 통해서 잘 작동하는 피쳐 엔지니어링을 하는 게 생각보다 매우 어렵고 감에 의존한 주먹구구식 피쳐 엔지니어링이 아니라 EDA 과정에서 얻은 명확한 인사이트를 구현하는 게 필요하다는 생각이 들었습니다.

마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?

이번 프로젝트에서는 EDA 를 진행하는 방법을 몰라서 유의미한 인사이트를 도출하지 못한 게 가장 아쉽습니다. 그리고 EDA 를 가장 잘한 1 위조의 발표를 보면서 기본적인 통계적 지식을 활용해서 명확한 근거에 기반하여 인사이트를 뽑아내는 과정이 어떻게 진행되는지 여러가지 실례를 찾아보고 익혀야겠다고 생각했습니다.

한계/교훈을 바탕으로 다음 프로젝트에서 스스로 새롭게 시도해볼 것은 무엇일까?

1 위를 한 6 조의 발표를 보고 EDA 에서 통찰을 건져내는 사례를 볼 수 있어서 좋았습니다. 1 위조의 프로젝트에서 두 가지가 가장 인상깊었습니다. 우선 첫째로 인상 깊었던 부분은 실험의 진행을 철저한 사전 분석에 기초해서 진행했다는 것입니다. 어떤 피쳐가 타겟값의 분포에 어떤 영향을 미치는지 비교적 간단한 데이터프레임 조작/시각화를 통해서 파악하는 과정이 프로젝트의 방향성에 엄청난 영향을 미친 것 같다는 생각이 들었습니다. 또한 이 과정에서 실험을 도와줄 수 있는 다양한 툴-Weight AND Bias-등을 적재적소에 사용하는 것도 인상적이었습니다. 둘째로 다음 프로젝트에서는 가설 설정 - 실험 - 가설 검증의 플로우를 반복해야겠다는 생각을 했습니다. 이번 프로젝트에서는 실험 결과를 꼼꼼히 기록하지 못했을 뿐만 아니라 프로젝트를 진행하는 전반에서 명확한 가설을 갖고 실험한 게 아니라 단지 현재보다 좋은 결과가 나오는지를 그때 그때 감에 따라서 판단했습니다. 그래서 프로젝트가 끝났지만 기록도 별로 남아 있지 않았습니다. 그 결과 특정 지점 이후에서는 체계적인 진전보다는 단지 운에 맡기는 양상을 보였습니다. 앞으로의 프로젝트는 실험 설정과 개선에 있어서 목표를 분명히하고 객관적인 근거를 갖추어 실험이 더 나은 방향으로 가는 데 도움이 되고 있음을 확인하는 방식으로 진행해야겠다는 생각이 듭니다.