**진행한 실험**

논문에서는 MLP 모델을 사용하는 경우에는 데이터를 embedding 하지 않고 사용하였기 때문에 일반적인 MLP 모델에 데이터를 전처리한 후, 해당 데이터를 입력하여 비교 진행하였습니다. 논문에서 제안한 WPFS(WPN과 SPN)의 경우, feature를 추출하거나 선택하는 방식으로, 두 네트워크의 입력값은 SVD, Dot Histogram 등을 사용하여 임베딩한 값이었습니다. 반면, 비교를 위해 MLP 모델에서는 Min-max scaling 전처리만 적용했으며, 임베딩 방식을 사용하지 않은 실험을 추가로 진행하였습니다(SVD, NMF, Dot Histogram을 사용한 실험 포함).

**실험 방식**

논문에서 제공한 깃허브 코드를 사용하려 했으나, 전처리 과정을 넣는 과정이 복잡하여 논문의 MLP 구조를 가져와서 기본적인 모델을 학습시켰습니다. 이를 통해 논문의 결과와 제공된 코드에서의 MLP 모델의 학습 및 검증 결과를 비교했습니다. 이후, 입력값을 다양한 임베딩 값으로 바꾸어가면서 실험을 진행했습니다. SVD, PCA, Dot Histogram 등의 feature selection 및 feature extraction 기법을 사용하여, SVD의 경우 몇 가지 feature를 선택하며 실험을 진행하고 결과를 분석했습니다.

**MLP 모델 정보**

논문에서 사용한 MLP 모델의 구조는 다음과 같습니다:

* **3개의 레이어**
  + 첫 번째 레이어: 100개의 뉴런
  + 두 번째 레이어: 100개의 뉴런
  + 세 번째 레이어: 10개의 뉴런
* 각 레이어는 LeakyReLU 활성화 함수를 사용합니다.
* 출력 레이어는 소프트맥스 활성화 함수를 사용하여 클래스 확률을 출력합니다.
* 모든 레이어는 배치 정규화(batch normalization)와 드롭아웃(dropout, p = 0.2)을 사용하여 과적합을 방지합니다.

**학습 및 검증**

* 손실 함수: 가중된 교차 엔트로피 손실 함수(weighted cross-entropy loss)
* 옵티마이저: AdamW
* 학습률: 초기 학습률 3e-3에서 3e-4로 선형 감소하는 스케줄러 사용
* 배치 크기: 8
* 에폭 수: 500
* 5-fold 교차 검증을 5번 반복하여 총 25번의 실험을 진행하였으며, 각 실험에서 학습 데이터의 10%를 검증 데이터로 사용했습니다.

**결과 및 분석**

본 논문의 결과보다 높은 정확도를 보이는 경우도 있고, 낮은 경우도 있었습니다. 하지만 모델 수 자체가 적고, 논문에서 정의한 문제가 데이터가 적을 때의 과적합 문제이기 때문에 논문의 방식을 따르기보다는 다양한 방식을 시도하여 최적의 모델을 찾는 것이 더 효율적일 것이라 생각됩니다. 만약 데이터의 암호화를 위해 모델 자체가 암호화된 경우처럼 결과를 확인할 수 없는 경우에는, 논문에서 제안한 모델의 효과에 대해 의문이 들 수 있습니다.

결과 표

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

**추후 진행 가능한 연구**

Dot Histogram 방식에서 성능이 비정상적으로 높게 나왔습니다. 이는 정확한 분석이 필요하지만, 시간 부족으로 진행하지 못했습니다. 다른 방식에 비해 비정상적으로 높기 때문에 신뢰할 만한 지표가 되지 못하는 것 같습니다. 또한, 논문에서 실험한 데이터셋 중 일부는 데이터를 구하지 못하여 실험하지 못했고, 인터넷에 공개된 5개의 데이터만으로 실험을 진행했습니다. 기타 다른 데이터가 가능할 경우, 다른 데이터에서 테스트를 통해 논문의 강건성을 확인하는 것도 중요한 과제가 될 것입니다.