

Data Augmentation Techniques

텍스트 증강

- 역번역: EN→KO→EN
- 동의어 치환: WordNet
- 패러프레이징: T5, GPT
- 임의 삽입/삭제

합성 데이터

- **GPT-4** 기반 생성
- 템플릿 기반 생성
- **SMOTE**: 소수 클래스
- **GAN**: 이미지 생성

증강 효과

데이터 다양성 증가, 과적합 방지, 소수 클래스 성능 향상, F1 score +5-15%