

059

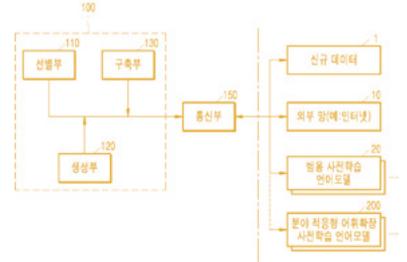
기술분류_ 인공지능

사전학습 언어모델 구축 및 어휘 확장

01 기술 개요

고성능 사전학습 언어 모델 구축 및 어휘 확장 기술

- 시간 및 비용의 상승을 최소화하면서 신규 데이터 기반의 어휘 확장을 통해 특정 분야를 위한 고성능 사전학습 언어 모델을 구축할 수 있는 기술을 실현하는 사전 학습 언어모델 구축장치 및 사전학습 언어모델의 어휘 확장 방법
- 구축된 범용 사전학습 언어모델에 추가 적용하기 위한 특정 어휘를 선별하고 그 어휘에 대한 임베딩 값을 생성하고, 범용 사전학습 언어 모델의 어휘 사전 및 임베딩 계층에 추가하여 특정 분야에 대한 분야 적응형 어휘 확장 사전학습 언어모델



[대표도면]

02 기술 차별성

언어 모델 구축에 소요되는 시간 및 비용 최소화

- 시간 및 비용의 상승을 최소화하기 위해 신규 데이터(예: 특정 분야의 데이터) 기반의 어휘 확장으로 언어모델을 구축함

집계연산을 통한 임베딩 값 생성

- 예측 어휘 집합에서 필터링된 예측어휘에 대한 임베딩 벡터값을 획득하고, 집계연산을 통해 특정 어휘에 대한 임베딩 값을 생성하기 위해 매체에 저장된 컴퓨터프로그램임

임베딩 계층을 용이하고 점증적으로 확장 가능

- 소규모의 신규 데이터와 적은 계산량 및 적은 비용으로 구체적인 기술 구성들을 실현 가능함
- 새로운 기술 출현이 빈번하고 이에 따른 새로운 용어 또한 빈번하게 발생하는 환경을 감안할 때, 어휘 사전 및 임베딩 계층을 용이하고 점증적으로 확장하여 사전학습 언어모델을 구축할 수 있는 요구사항을 충족시키는 효과가 있음

03 기술 키워드

사전학습 언어모델, 인공지능 딥러닝, 어휘 확장

04 기술의 TRL 단계



059

기술분류_ 인공지능

사전학습 언어모델 구축 및 어휘 확장

05 사업화 포인트

딥러닝은 사람의 입력에 의존하지 않고 작업을 수행할수 있는 능력과 여러 산업분야에서 클라우드 기반 기술이 빠르게 채택되면서 딥러닝 산업의 성장이 가속화되고 있음

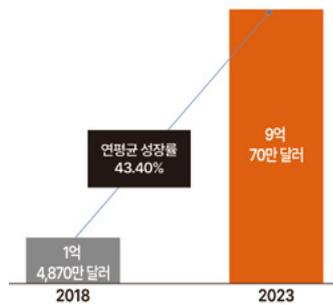
06 활용 분야 및 시장 규모

활용 분야

사전학습 언어모델, 딥러닝

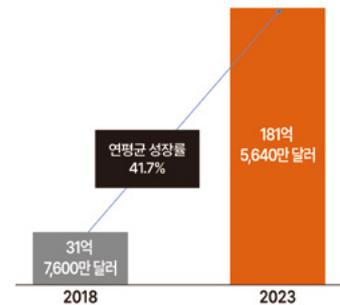
시장 규모 및 전망

[국내 딥러닝 시장]



(출처: 연구개발특구진흥재단 딥러닝 시장 보고서)

[국외 딥러닝 시장]



(출처: 연구개발특구진흥재단 딥러닝 시장 보고서)

07 지식재산권 현황

권리현황

특허명	사전학습 언어모델 구축장치 및 사전학습 언어모델의 어휘 확장 방법
출원번호	10-2022-0077942
권리자	한국과학기술정보연구원
관리기관	한국과학기술정보연구원
담당자	주용하
문의처	042-869-0977