

개념 관계 추출을 활용한 LEVEL 기반 동적 기사 추천 시스템

추천 시스템 2조 | 박예슬 안유민 유현선 황서현

BITAmin 25-1 프로젝트 발표

주제 선정 배경

Pipeline

모델 구축

모델 시연

결론

데이터 수집 및 구축

- 기사 데이터
- 논문 데이터
- 관계 데이터

- 이론적 내용
 - 관계 추출
 - 강화학습
- 모델 구조
 - DyGIE++ 모델
 - Embedding 모델
 - PPO 알고리즘
- 모델 적용

- 웹 화면 구성

- 인사이트 도출
- 개선 사항 및 한계점

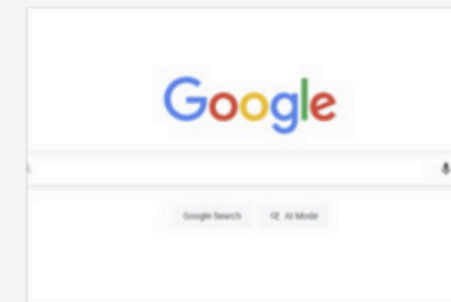
기존 뉴스 레터의 한계점



- 정기적으로 주 1회 랜덤으로 제공되는 뉴스 레터
- 개인 관심사 반영 X

기존 추천 시스템의 한계점

키워드 관련기사



구글, 검색 페이지에 'AI 모드' 버튼 추가..."챗봇형 검색 본격 테스트"



[5월12일] "구글이 오픈AI에 뒤쳐진 것은 검색 때문...광고 수익 집"



구글 "검색 트래픽 안 떨어져"...전문가는 "본격 AI 도전 시작돼"

- 유사도 기반 추천 시스템의 경우, 사용자의 지식 발전도를 판별하는 데 아쉬움이 있음
- 사용자의 업데이트 되는 지식 상태 반영 필요

기존 뉴스 레터의 한계점

기존 추천 시스템의 한계점

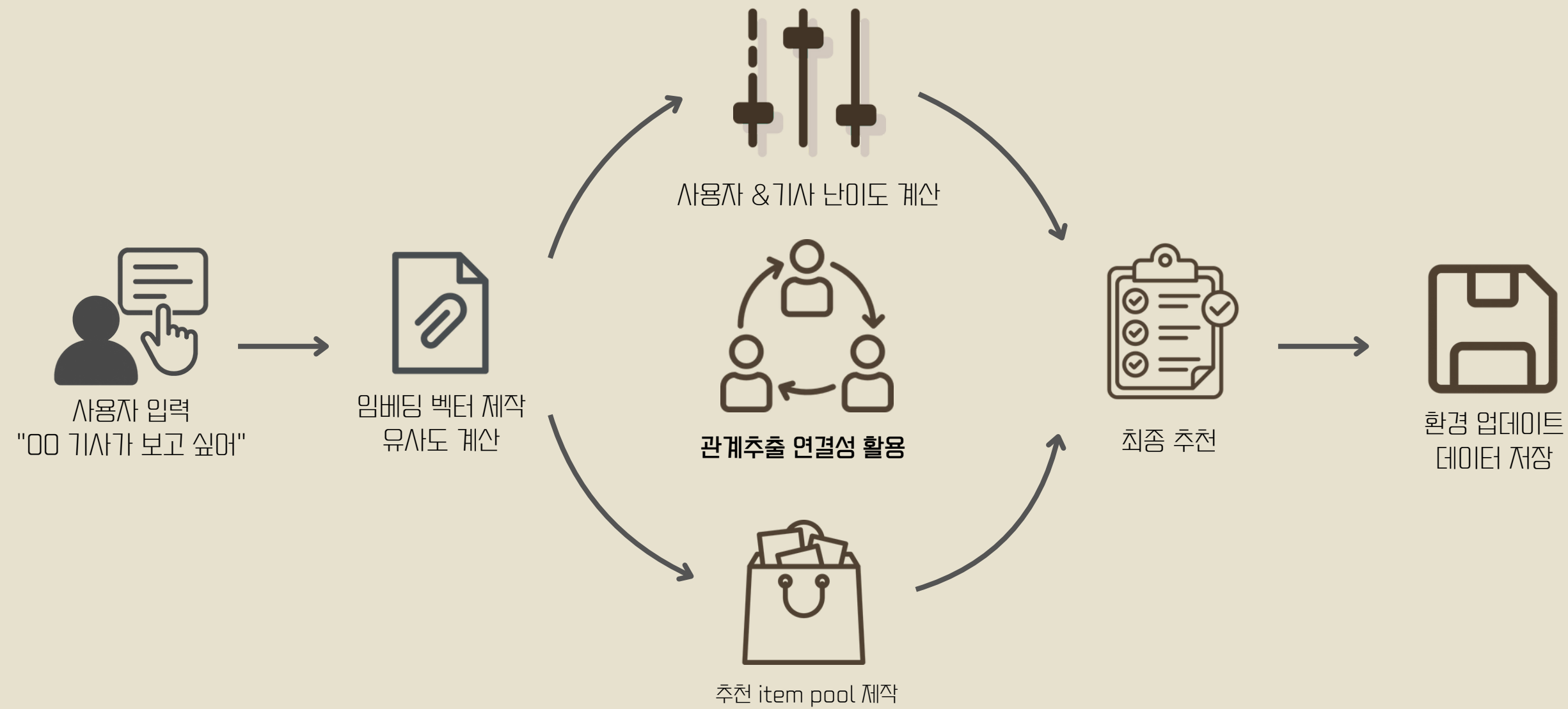
개인 지식 발전도 + 사용자 맞춤형을 모두 충족하는 “개념 관계 추출을 활용한 level 기반 동적 기사 추천 시스템”

- 정기적으로 주 1회 랜덤으로 제공되는 뉴스 레터
- 개인 관심사 반영 X

- 유사도 기반 추천 시스템의 경우,
사용자의 지식 발전도를 판별하는 데 아쉬움이 있음
- 사용자의 업데이트 되는 지식 상태 반영 필요

Pipeline

SYSTEM



TRAIN



state



action



reward



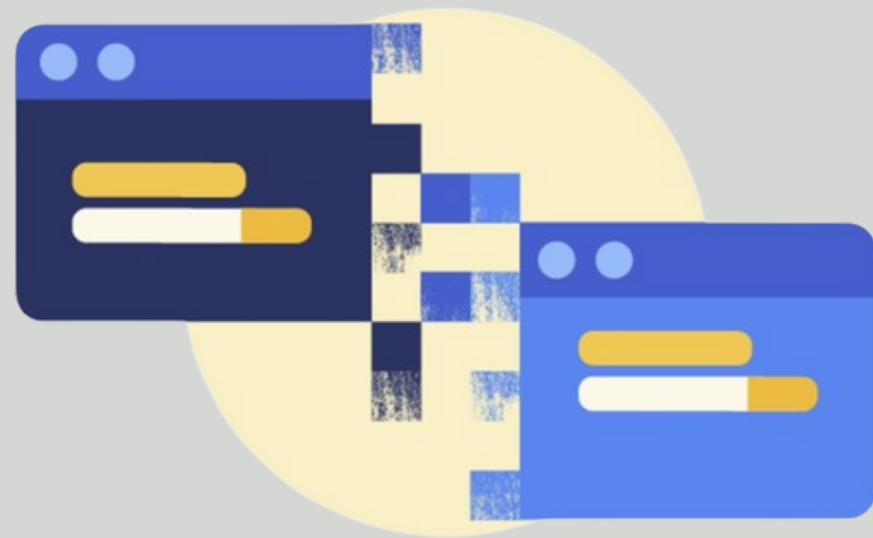
update

데이터 수집: 기사 데이터



AI TIMES(인공지능 신문) 사이트에서
기사 카테고리가 머신러닝, 딥러닝, 뉴럴 네트워크, 자연어, 음성, 이미지이고,
발행 연도가 2020년 이후인 기사들 크롤링

title	date	category	content
AI, 남극 빙하 이동의 비밀 을 밝히다...스탠퍼드대 라이 교수팀, "해수면 상승 예측 정확도 높여"	2025.03.14	Machine Learning	기후 변화로 인해 남극의 빙하가 빠르게 녹으며 해수면 상승을 가속화하고 있다. 남극 대륙에는 전 세계 해수면을 최대 58m 상승시킬 수 있는 엄청난 양의 얼음 ...
인공지능이 3D 프린팅 티타늄 합금 성능 향상... 항공·조선· 의료 산업 혁신 기대	2025.03.10	Machine Learning	美 미국 존스홉킨스응용물리연구소(Johns Hopkins Applied Physics Laboratory)와 존스 홉킨스 위팅 공과 대학(Johns Hopkins Whiting School of Engineering) 연구진이 인공지능(AI)을 활용해 3D 프린팅 티타늄 합금의 성능을 ...



Semantic Scholar API

[논문 키워드 기반 검색]

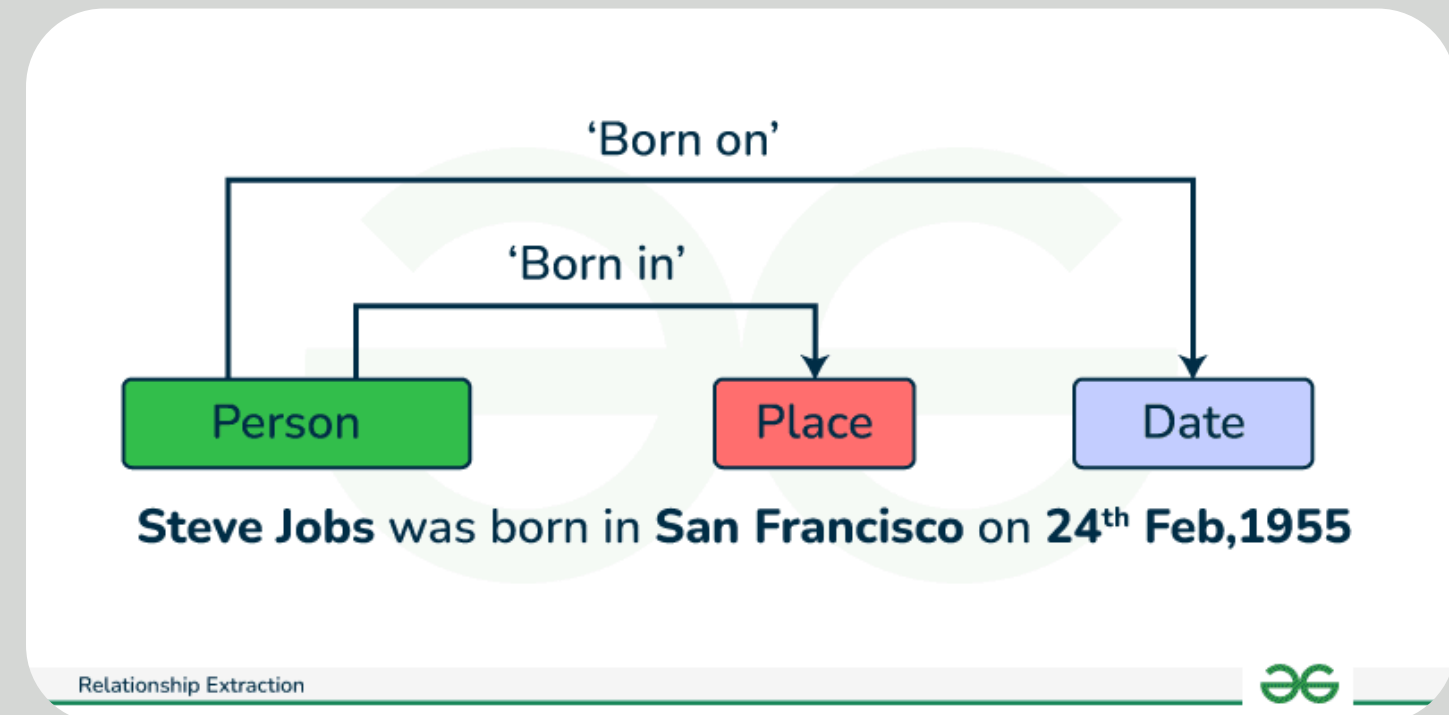
관계 추출용 데이터 수집
1416개

1. recommendation system 등의 주요한 주제 15개 검색
2. 초록 문서 내부 지시어 해소 후 저장

title	year	category	abstract
Efficient Memory Management for Large Language Model...	2023	Computer Science	High throughput serving of large language models (LLMs) requires batching sufficiently many requests at a time. However, existing systems struggle because.....

관계 추출이란?

- 문장의 단어(Entity)에 대한 속성과 관계를 예측하는 task
- <주어, 관계, 목적어>의 tripple 형태로 두 개체 간의 관계(relation)을 추출



IT 전문 용어들 간 그래프 구축
→ 상의어, 하의어 relation 기반 난이도 계산 용이

DyGIE++

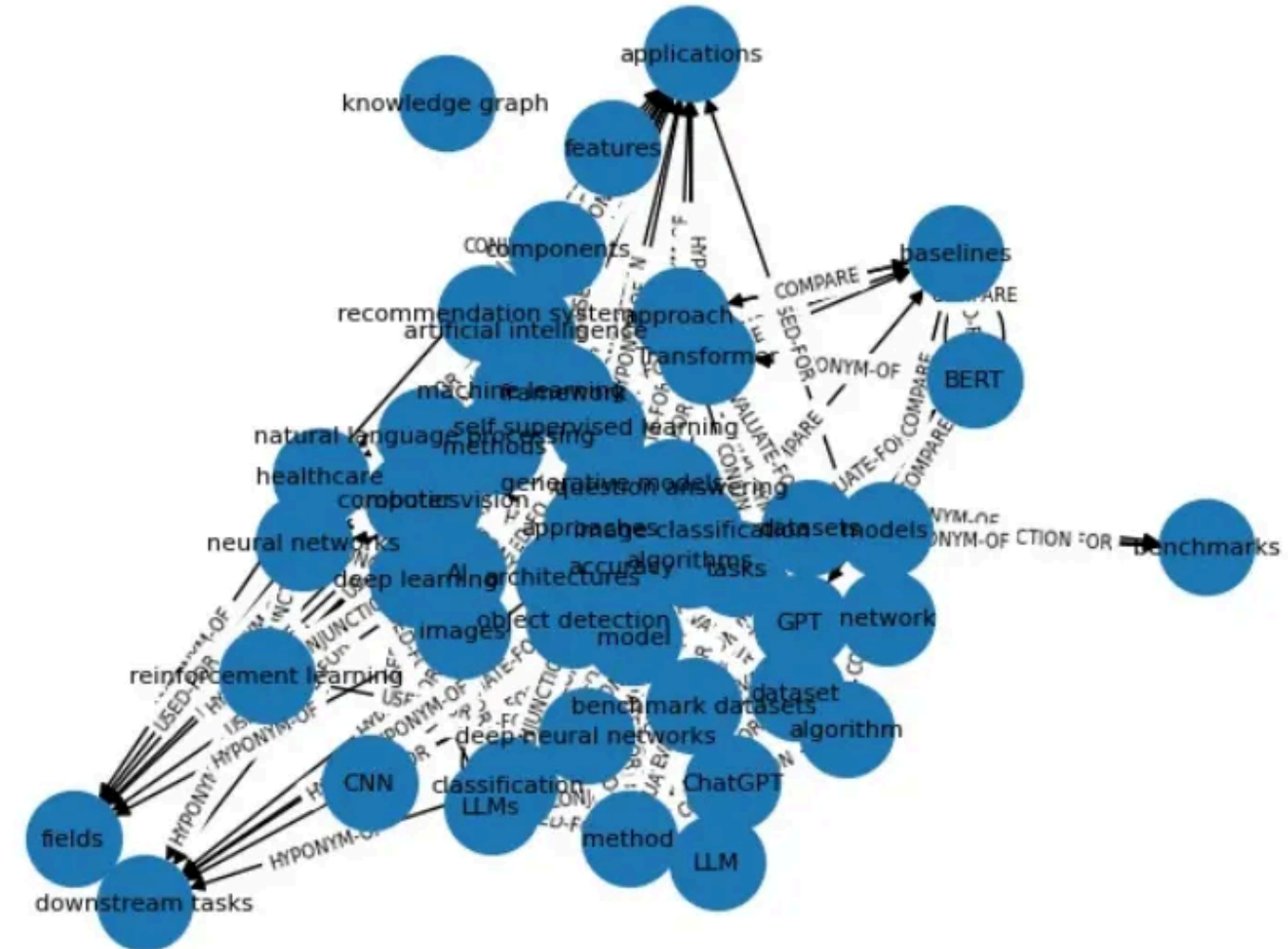
- 정보 추출을 위한 모델 중 하나로, 특히 문서 단위에서 개체명 인식(NER), 관계 추출(RE), 이벤트 추출(EE) 등을 동시에 처리하는 **멀티태스크 모델**
- 문서 전체를 고려해서 정보를 뽑아내기에 단순한 문장 단위 RE보다 **더 정밀하고 넓은 관계**를 잡을 수 있음

토큰 → Span 추출 → Span 간 관계 추론 → 결과 예측

데이터 구축: 관계 데이터

총 16439개 데이터 추출

subject	mapped_relation	object
LOGISTIC REGRESSION	HYPONYM-OF	SUPERVISED LEARNING
LINEAR REGRESSION	HYPONYM-OF	SUPERVISED LEARNING
DECISION TREE	HYPONYM-OF	SUPERVISED LEARNING
RANDOM FOREST	HYPONYM-OF	SUPERVISED LEARNING
SUPPORT VECTOR MACHINE	HYPONYM-OF	SUPERVISED LEARNING
K-NEAREST NEIGHBORS	HYPONYM-OF	SUPERVISED LEARNING
GRADIENT BOOSTING	HYPONYM-OF	SUPERVISED LEARNING
NAIVE BAYES	HYPONYM-OF	SUPERVISED LEARNING
SUPERVISED LEARNING	USED-FOR	CLASSIFICATION
SUPERVISED LEARNING	USED-FOR	REGRESSION
SUPERVISED LEARNING	FEATURE-OF	LABELED DATA
UNSUPERVISED LEARNING	HYPONYM-OF	CLUSTERING
UNSUPERVISED LEARNING	USED-FOR	DIMENSIONALITY REDUCTION
UNSUPERVISED LEARNING	FEATURE-OF	UNLABELED DATA



데이터 정제: 관계 데이터

Preprocessing

Word Extraction
Korean Mapping

약어 추출

DL → DEEP NEURAL NETWORK

분사형 변환

LEARN → LEARNING

띄어쓰기 & 오타

PSEUDOLABELS
→ PSEUDO LABELS

단수형 변환

CNNS → CNN

모호한 복합어

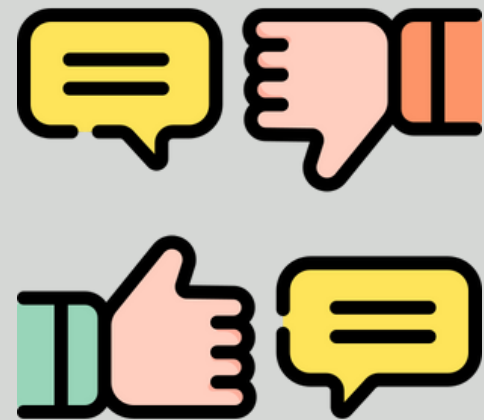
AI MODEL, AI MODELING
→ AI MODEL

AND 치환

VISION AND LANGUAGE MODEL
→ LANGUAGE VISION MODEL

에이전트가 보상을 최대화 하기 위해 스스로 시행착오를 겪으며 **최적의 행동**을 학습

Feedback



사용자의 다양한 피드백을
보상으로 환산하여
모델이 스스로 추천 개선

장기적 Reward 개선



사용자가 지금은 특정 난이도의
기사를 이해하지 못하더라도,
장기적인 향상 학습에
도움이 되는 경로 유도 가능

개인화 State - Action

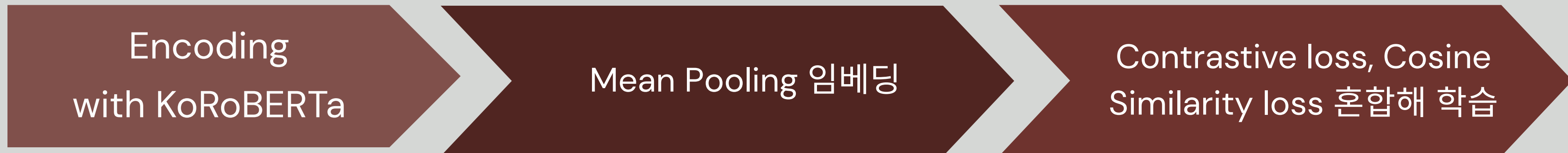


사용자의 수준에 따라
초급 → 중급 → 고급으로
성장하는 개인 맞춤 학습 경로 수립

모델 구조: Embedding 모델

BM-K/KoSimCSE-roberta-multitask 모델

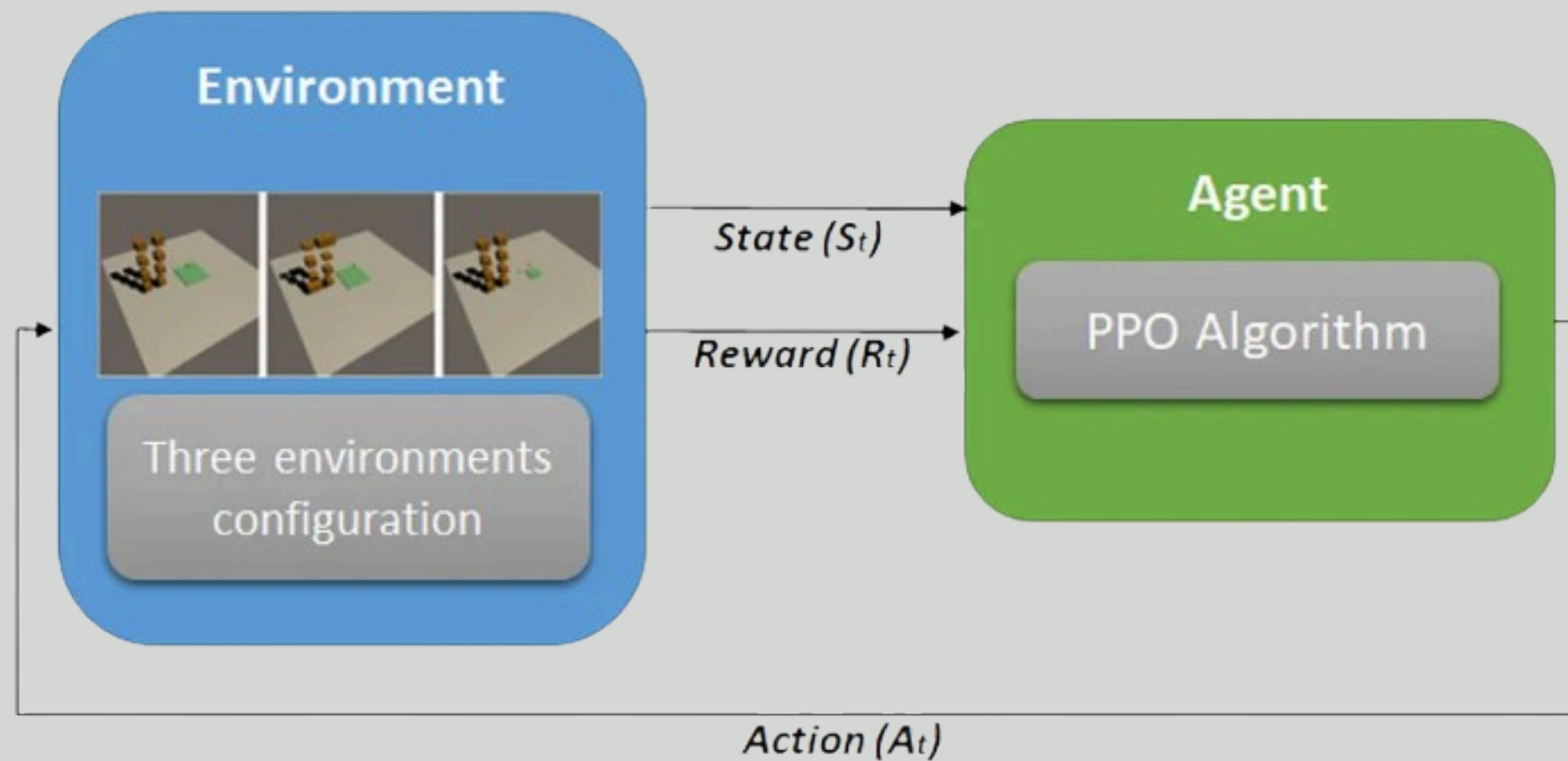
한국어 자연어 처리에서 문장 간 유사도 분석이나 검색, 분류 등에 매우 효과적인 Embedding 모델로 평가



STS, NLI, 문장 유사도 ranking 등 멀티태스크 학습
→ 단순 유사도 + 세밀한 의미 차이 + 문맥 흐름까지 반영한 Embedding 모델

System Embedding 모델 채택

모델 구조: PPO 알고리즘



Proximal Policy Optimization Algorithms

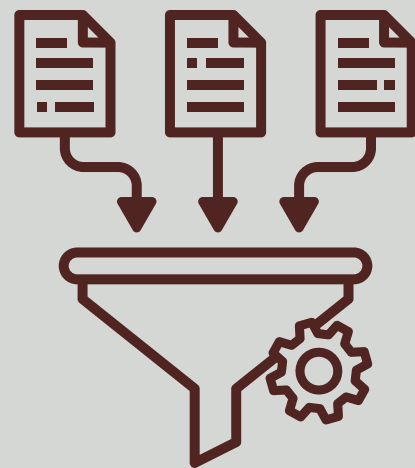
- 강화학습에서 사용되는 정책 기반 알고리즘
- 에이전트가 환경과 상호작용
→ **보상을 최대화하는 정책을 학습**
- 학습 안전성 보장 가능

유사도 기반
가상 User 선호도 제작



OFFLINE 정책 일부 차용

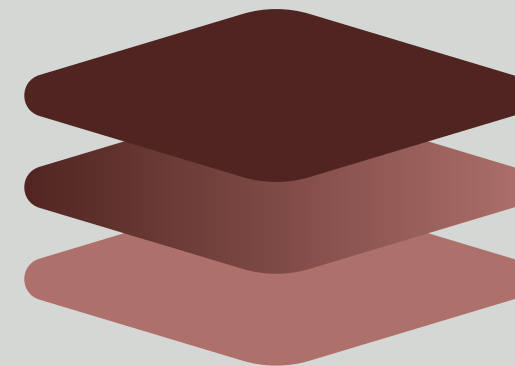
Concept 01



동적 난이도 추정

- 관계 데이터를 통한 user 별 입력 난이도 산정
- User state를 기준으로 기사별 난이도 추정

Concept 02



보상 설계

- 선호도, 유사도, 난이도 복수 누적 보상 조정 (가중치 반영)
- 이전 난이도를 반영하여 세부적인 업데이트 수행

모델 시연: 웹 화면 구성



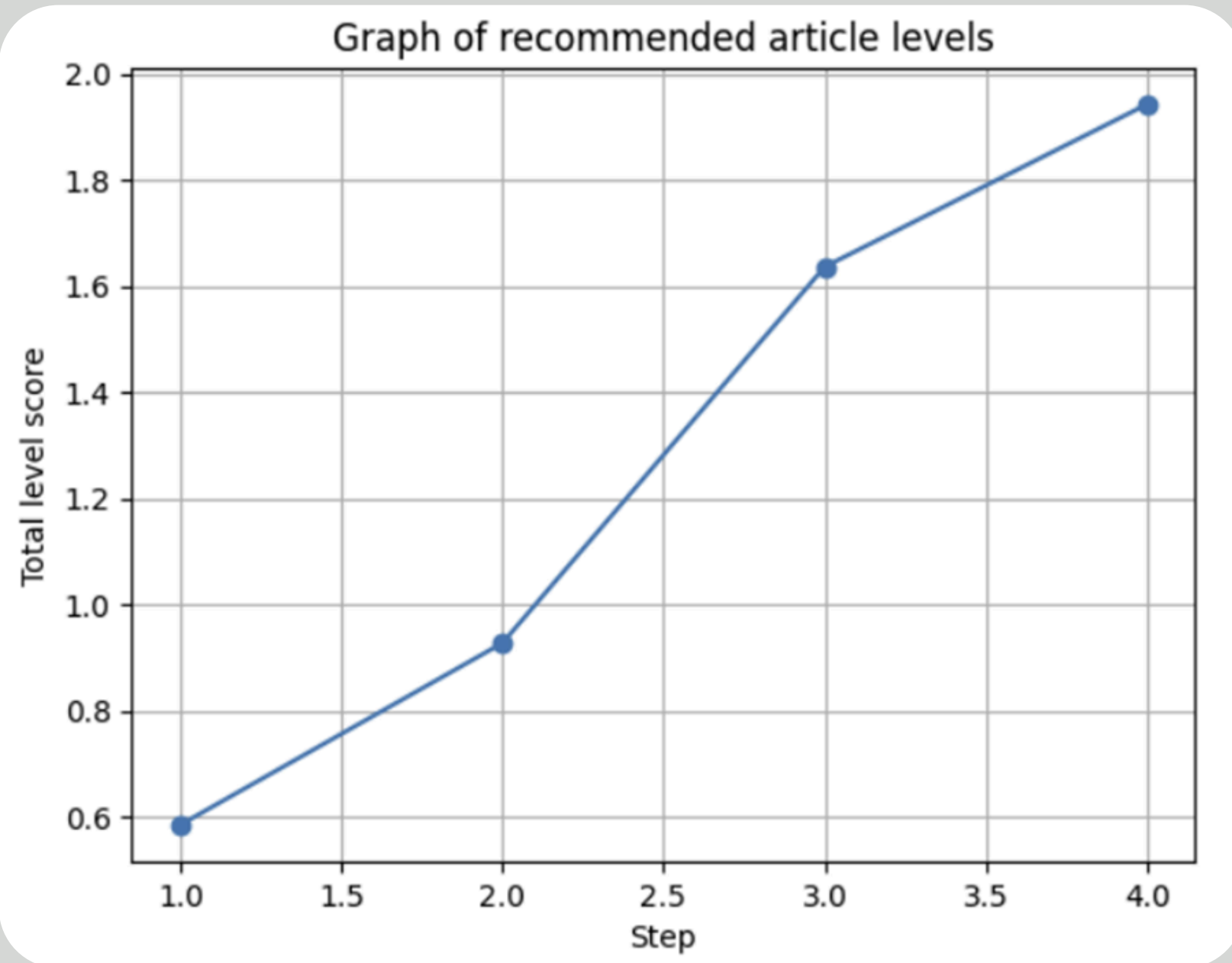
FastAPI

- Python 기반의 웹 프레임워크
- HTML과 합쳐서 로컬 웹 테스트 수행

시연 영상



결론: 인사이트 도출



main topic

사용자 입력과 개념 추출 데이터 관계성을
바탕으로 기사 난이도 측정
&
강화학습 기반 모델을 도입하여
사용자에게 점진적으로 난이도가 상승하는 기사 제공

특정 라벨링 없이
사용자 맞춤형 난이도 조절 알고리즘 설계

강화학습 기반 추천의 유효성 확인

결론: 개선 사항 및 한계점

Improvements

DB 연결

입력 약어 처리

보상 함수 최적 가중치

Limitation

서버 및 GPU 사용량 문제

노드 표현성 문제

THANK you

추천시스템 2조
박예슬 안유민 유현선 황서현