


도메인 특화 텍스트 및 멀티모달 임베딩 기술

숭실대학교 AI융합학부 ● 박건우 교수

기술개요

-  본 기술은 인용구 의미 표현이 가능하도록 하여 제목에 날조된 인용구를 사용하는 콘텐츠의 탐지력을 향상시키는 것은 물론, 임베딩 모델 학습 시 필요한 라벨 데이터를 생략할 수 있어 라벨 데이터에 필요한 비용을 감소시킬 수 있는 대조학습 기반의 날조된 뉴스 인용구 탐지 방법에 관한 것임

기술성숙도(TRL)

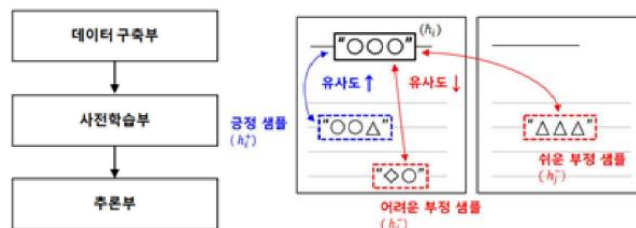
1	2	3	4	5	6	7	8	9
기초연구		실험		시작품		실용화		사업화

기술 개발 배경

- 웹 및 모바일 환경에서 많은 정보가 짧은 제목 텍스트와 함께 공유되며 해당 정보가 본문을 대표하지 않을 경우 공유되는 정보의 내용을 오도할 수 있다는 문제가 있으며, 일 예로 제목에 선정적인 내용을 사용하는 낚시성 기사가 있음
- 이를 위해 관련성이 떨어지는 정보를 탐지하기 위한 기술이 제안되고 있지만, 디지털 환경에서는 대부분 제목과 요약, 이미지에 기반해 링크가 공유되며 링크를 클릭하지 않는다면 진위 여부를 판단하기 어렵다는 단점이 있음

기술 차별성

- 텍스트를 대표하지 않는 이미지를 사용하는 콘텐츠를 탐지할 수 있는 시각-언어 의미 표현이 가능한 학습 방법은 물론, 임베딩 모델 학습 시 정답 레이블이 없는 데이터를 이용하여 레이블링 비용 절감할 수 있음
- 인용구 의미 표현이 가능하도록 하여 제목에 날조된 인용구를 사용하는 콘텐츠의 탐지력을 향상시키는 것은 물론, 임베딩 모델 학습 시 필요한 라벨 데이터를 생략할 수 있어 라벨 데이터에 필요한 비용을 감소시킬 수 있음



도메인 특화 텍스트 및 멀티모달 임베딩 기술

숭실대학교 AI융합학부 ● 박건우 교수

기술 활용 분야

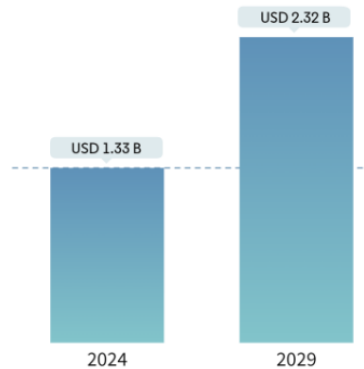
- 데이터 마이닝 분야
 - 기업 브랜딩 및 마케팅
 - 저작권 및 무결성 확인
- AI 모델 및 플랫폼 개발 분야
 - 멀티모달 임베딩 SaaS 플랫폼
 - AI 콘텐츠 필터링 솔루션



시장동향/개발 현황

- 관련 시장 동향
 - mordorintelligence의 보고서에 따르면, 데이터 마이닝 시장 규모는 2024년에 13억 3천만 달러로 추산되며, 2029년까지 23억 2천만 달러에 도달할 것으로 예상되며, 예측 기간 동안 연평균 성장률은 11.8%를 기록할 것으로 예상됨
- 개발 동향
 - 데이터 마이닝과 딥러닝 모델의 통합으로, 복잡한 비정형 데이터(텍스트, 이미지, 음성 등)에서도 높은 정확도의 패턴을 탐지 가능함
 - 텍스트, 이미지, 음성 등 다양한 유형의 데이터를 함께 분석하는 멀티모달 분석 기술이 주목받고 있음

Data Mining Market
Market Size in USD Billion
CAGR 11.80%



Source : Mordor Intelligence

[글로벌 사이버 보안 시장 전망(A)]
(출처: mordorintelligence)

지식재산권 현황

No	특허명	출원번호	등록번호
1	대조학습 기반의 날조된 뉴스 인용구 탐지 방법 및 장치	KR 10-2024-0030153	-
2	뉴스 텍스트와 관련성이 떨어지는 이미지 탐지를 위한 이미 지-텍스트 사전학습 방법 및 탐지 장치	KR 10-2024-0030158	-