30118 이수민 · 30219 윤영서

임베딩 모델을 통한 알고리즘 편향도 평가 서비스 개발

목차

개발 동기

알고리즘

현재 SNS, 쇼핑몰 등 다양한 서비스에서

사용자의 취향에 맞춰 콘텐츠를 추천해줌

사용자가 자신에게 필요한 정보를 검색하는 등

추가적인 노력 없이 알맞은 정보를 얻을 수 있음

필터버블

사용자가 관심을 가지는 주제의 콘텐츠만 추천함

사용자는 다양한 정보를 접하지 못하고

편향된 정보만을 제공받게 됨

확증 편향을 강화시키고 건전한 사고를 해침

연구 목적

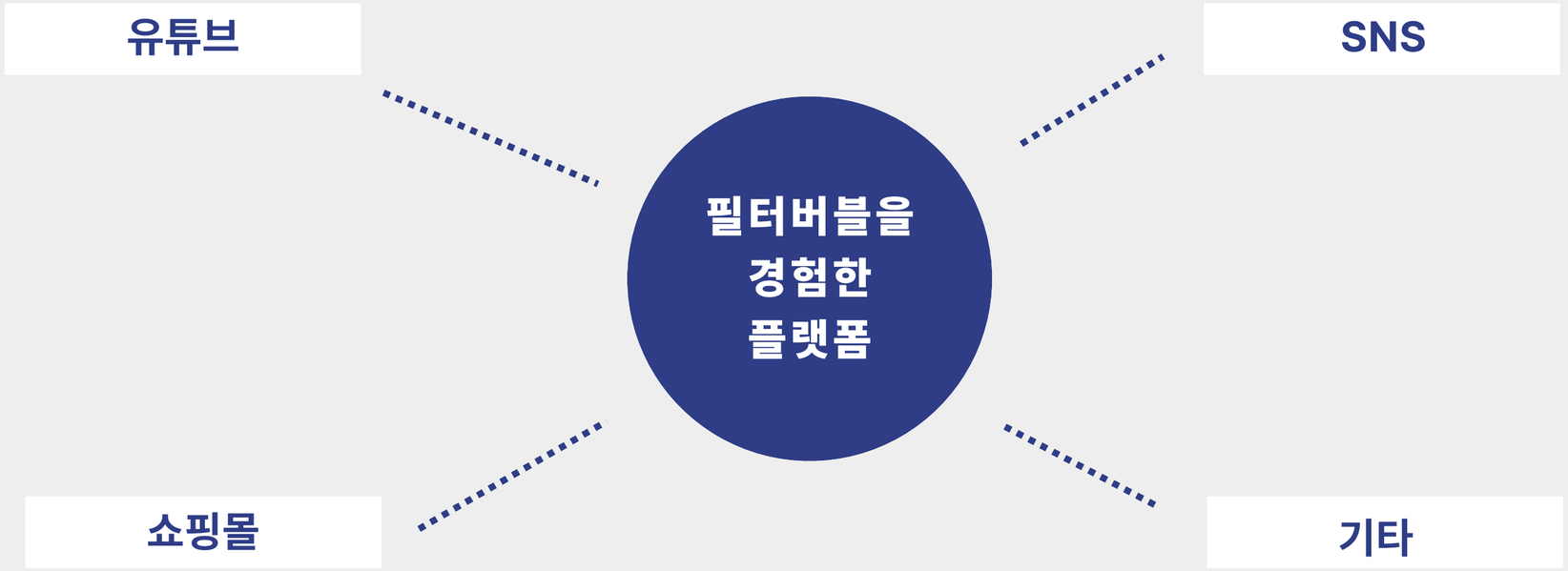
이와 같은 문제를 해결하기 위해

추천 알고리즘을 추적하여 편향도를 측정하는

서비스를 개발하고자 함

개요

교내 학생 18명에게 설문 조사를 실시한 결과 3명이 필터 버블에 대해 알고있다고 답하였고
전원 모두 필터 버블을 경험한 적이 있다고 답하였다. 이들 중 대부분이 유튜브에서 필터 버블을 경험하였고
SNS, 쇼핑몰 순으로 경험한 플랫폼이 많았다. 응답 중 필터 버블이 가장 많이 경험된
유튜브의 알고리즘 편향도 평가 서비스를 개발한다



서비스 작동 원리

임베딩

임베딩은 인간이 사용하는 자연어를
고차원 공간에 벡터로 표현하는 것
(본 서비스에서는 3072차원)

유사도 측정

각 명사와 주제들의 유사도의 평균을 내어
해당 제목을 가진 영상이 어떤 주제를 다루고 있는지 판단
본 과정을 수차례 반복하여 사용자의 추천
알고리즘이 어떤 주제의 영상을 어떤 비중으로 추천하는지 측정

제목 임베딩

유튜브 추천 알고리즘을 통해서 콘텐츠의 제목을 불러온다.
제목에서 명사만을 추출한 뒤에 임베딩
추출된 명사를 미리 임베딩 시켜둔
정치, 동물등의 주제와 비교하여 유사도를 측정
비교는 벡터의 각도를 이용한 코사인 유사도를 통해 진행

결과

이때 일정 수준 이상으로 한 주제가 추천된다면
이는 알고리즘이 편향되었다고 간주함

향후 계획

서비스를 사용자들이 사용할 수 있도록 웹페이지로 공개

유튜브의 경우 제목 뿐 아니라 썸네일, 설명, 태그 등 사용자가 콘텐츠의 주제를 짐작할 수 있는 요소가 존재하므로

이러한 요소들을 고려한다면 정확도가 오를 것

따라서 설명과 태그, 제목과 같은 텍스트 형식 정보를 취합하여 판단 할 수 있도록 개선

이미지 모델을 활용하여 썸네일 또한 다뤄 서비스 개선을 목표로 함

유튜브 제목

유튜브 콘텐츠를 직관적으로 설명하는 요소인
제목을 통해 1차적으로 편향을 알아봄

썸네일, 설명, 태그 등

서비스의 정확도를 올리기 위해 다양한 요소들을
고려할 수 있도록 개선함

감사합니다