

실시간 IPTV 방송스트림 기반 멀티모달 상품정보 추론 및 스코어링 시스템 개발

배주한, 허재호, 주현철, 나태영
SK 텔레콤

{juhan.bae, jaeho.hur, hc.joo, taeyoung.na}@sk.com

Multi-Modal Product Information Inference and Scoring System Based on IPTV Real-Time Broadcast Streams

Juhan Bae, Jaeho Hur, Hyunchul Joo, Taeyoung Na
SK Telecom

요 약

본 논문은 IPTV 실시간 방송 스트림을 분석하여 상품 정보를 추론하고 신뢰도 점수 기반으로 스코어링하는 멀티모달 시스템을 제안한다. 영상, 자막, 화면 내 텍스트를 통합 분석해 상품을 실시간 인식하고, 메타데이터를 구조화하여 전자상거래 플랫폼과 연동한다. 이를 통해 시청자는 방송 중 등장한 상품을 즉시 탐색·구매할 수 있으며, 본 연구는 스마트 TV 기반 실시간 커머스의 새로운 가능성을 제시한다.

I. 서 론

스트리밍 미디어 소비가 증가하면서 방송 콘텐츠와 상업적 정보의 융합을 통한 새로운 소비자 경험이 주목받고 있다. IPTV 실시간 방송에서는 시청 중 상품 정보를 즉시 탐지하고 제공하는 기술 수요가 급증하며, 이는 인터랙티브 커머스(T-Commerce)로 발전하고 있다.

기존 광고 기반 커머스 모델은 사용자의 수동적 반응에 의존하거나 사전 정의된 상품 정보를 전달하는 한계를 가진다. Google YouTube 는 콘텐츠 제작자의 사전 등록이 필요하여 VoD 콘텐츠에 한정되고, Amazon Live 는 홈쇼핑 형태의 라이브 스트리밍 서비스를 제공하며, Facebook/Instagram Shopping 은 SNS 기반 라이브 방송 중 즉시 구매 기능을 구현하고 있다. 국내 서비스의 경우, SK Broadband 는 VoD 서비스 내 수작업으로 등록된 PPL 상품 정보 기반의 구매 QR 코드를 제공하고 있으나 실시간 방송을 지원하지 않는다. 네이버 쇼핑라이브는 스마트스토어 입점 업체 대상 라이브 커머스 톨을 지원하여 실시간 방송을 통한 상품 소개 및 즉시 구매 연계 서비스를 제공하고 있다.

그러나 상기 서비스들은 사전 정의 의존성, 실시간 방송 적용 제한, 단일 모달리티 분석이라는 공통적인 한계를 지닌다. 기 구축된 상품 정보에 의존하여 새로운 상품의 실시간 반응이 어렵고, 대부분 VoD 콘텐츠 대상이며 실시간 연동이 불가능하다. 또한 영상, 자막, 화면 내 텍스트 등 다양한 소스의 통합 분석 기능이 제한적이다.

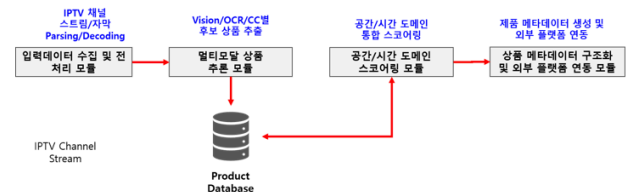
본 연구는 방송 중 등장하는 상품을 실시간 인식하여 즉시 추천하고 구매 연계하는 상호작용 기반 커머스 시스템을 구현하고자 한다. 관련 특허 기술로는 라이브 영상 스트림과 쇼핑 인터페이스 연동[1], 실시간 방송영상과 부가정보 동기화[2], IPTV 기반 인터랙티브

쇼핑[3] 등이 있으나, 멀티모달 데이터의 통합 분석과 실시간 상품 추론에는 한계가 있다.

II. 본론

1. 제안하는 시스템 구조

본 논문에서 제안하는 시스템은 입력 데이터 수집 및 전처리 모듈, 멀티모달 상품 추론 모듈, 시간도메인 스코어링 모듈, 상품 메타데이터 구조화 및 외부 플랫폼 연동 모듈의 4개 핵심 모듈로 구성된다.



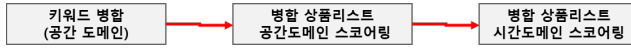
<그림 1. 전체 시스템 구조도>

입력 데이터 수집 및 전처리 모듈은 IPTV 방송 스트림으로부터 실시간 콘텐츠 데이터를 수집하고 멀티모달 분석을 위한 입력 형태로 전처리를 수행한다. 각 모달 데이터는 시간 동기화를 기준으로 병렬 수집되며, 통합 분석을 위해 멀티모달 상품 추론 모듈로 전달된다. 영상 프레임은 IPTV 스트림 파싱 및 디코딩 후 일정 주기로 키 프레임을 추출하며, 자막 정보는 IPTV 방송 스트림에서 제공되는 자막 데이터를 실시간으로 추출한다. 화면 내 텍스트는 방송 화면의 그래픽 문자, 오버레이 텍스트 자막 외 시각적 문자를 OCR(Optical Character Recognition) 기술을 활용하여 추출한다.

멀티모달 상품 추론 모듈은 전처리된 멀티모달 데이터를 입력으로 하여 방송 콘텐츠에 노출된 상품을 실시간 인식하고 상품 속성 정보를 추론한다. Vision 추론에서는 이미지 내 상품 탐지 및 상품 키워드, 관련 상품 크롭 영역을 추출하고, Caption 추론에서는 자막에서 텍스트 기반 상품 키워드를 추출한다. OCR 추론에서는 화면 내 텍스트 분석을 통한 텍스트 기반 상품 키워드를 추출한다. 최종 추론된 상품 정보는 스코어링 모듈로 전달된다.

2. 공간/시간 스코어링 알고리즘

시간공간 스코어링 모듈은 모달리티별 상품 추론 결과에 대해 통합 분석된 추천 상품 리스트 및 상품별 스코어를 산출한다. Vision, OCR, Caption 추론 결과인 상품 키워드를 대상으로 현재 시점 기준 최근에 집중적으로 여러 모달리티에서 반복해서 검출된 상품 키워드에 높은 스코어를 할당하며 아래와 같은 순서로 진행된다.



<그림 2. 시간공간 스코어링>

(1) 키워드 병합

공간도메인에서 상품 키워드 병합 과정은 다음과 같이 수행된다. 초기 모달리티별 기본 스코어($Score_{base}^{vision}$: $Score_{base}^{ocr}$: $Score_{base}^{cc}$)는 공간 도메인 상 각 모달리티별 중요도를 의미하며, 사용자 편의 및 콘텐츠 특징에 따라 변경 가능하며 시점 t 에서 모달리티별 추출 상품의 집합은 아래와 같이 정의한다.

$$ITEM_{method}(t) = \{item_{method}^1(t), item_{method}^2(t), \dots, item_{method}^{max_method}(t)\}$$

시점 t 기준 공간도메인 중복상품 제거를 위해 병합 리스트($ITEM_{merge}(t)$)를 만들며 모달리티 간 상품 키워드들의 텍스트 인코딩된 값의 코사인 유사도(Cosine Similarity) 값이 임계치(Th) 이상일 경우 동일 상품 키워드라 간주한다. 또한, 다수 모달리티에서 추출된 상품에 높은 가중치를 주기 위하여 모달리티 카운트($modality_count_l(t)$)를 계산하며 아래와 같이 표현된다.

$$modality_count_l(t) = \sum_{vi} \Delta_{vision}^{li}(t) + \sum_{vj} \Delta_{ocr}^{lj}(t) + \sum_{vk} \Delta_{cc}^{lk}(t)$$

$$= \begin{cases} \Delta_{method}^{lm}(t) & \text{if } sim(E^T(item_{merge}^l), E^T(item_{method \in \{i,j,k\}}^{m \in \{i,j,k\}})) > Th \\ 0 & \text{otherwise} \end{cases}$$

l = Index of merged product list
 i = Index of product list from vision
 j = Index of product list from OCR
 k = Index of product list from CC
 E^T = Text Encoder

(2) 병합상품리스트 공간도메인 스코어링

$modality_score_l(t)$ 는 시점 t 기준으로 병합리스트의 상품 l 이 다수의 모달리티에서 많이 검출되면 될수록 높아지는 값으로 모달리티별 기본점수($Score_{base}^{method}$)에 모달리티별 검출유무($\sum_{vi} \Delta_{method}^{li}(t)$)와 모달리티 검출 가중치($W_{spatial}^{modality_count_l(t)}$)의 곱으로 표현된다.

$$modality_score_l(t) = \max \left(Score_{base}^{vision} \times \sum_{vi} \Delta_{vision}^{li}(t), Score_{base}^{ocr} \times \sum_{vj} \Delta_{ocr}^{lj}(t), Score_{base}^{cc} \times \sum_{vk} \Delta_{cc}^{lk}(t) \right) \times W_{spatial}^{modality_count_l(t)}$$

$$W_{spatial}^{modality_count_l(t)} = \begin{cases} W_{spatial}^1 & \text{if } modality_count_l(t) = 1 \\ W_{spatial}^2 & \text{if } modality_count_l(t) = 2 \\ W_{spatial}^3 & \text{if } modality_count_l(t) = 3 \end{cases}$$

(3) 병합상품리스트 시간도메인 스코어링

$W_{temporal}^x$ 는 시점(t)기준으로 이전 $M-1$ 개의 프레임 을 검색범위, 검색범위내 프레임 위치를 x 라 가정 할때 현재시점에 가까울수록 1에 가까운 값을 가지며 아래와 같이 표현된다.

$$W_{temporal}^x = W_{temporal}^{base} + ((1 - W_{temporal}^{base}) \times \frac{x}{M})$$

따라서 t 시점 기준 병합상품 l 의 최종스코어($final_score_l(t)$)는 아래와 같이 표현된다.

$$final_score_l(t = M) = \sum_{1 \leq x \leq M-1} \left(W_{temporal}^x \times modality_score_l(x) \times \sum_{vj} \Delta_{temporal}^{lj}(x) \right) + W_{temporal}^M \times modality_score_l(M)$$

3. 상품 메타데이터 구조화 및 외부 플랫폼 연동

상품 메타데이터 구조화 및 전자상거래 플랫폼 연계 모듈은 스코어링 모듈에서 선출된 각 상품의 메타데이터를 정형화하고 외부 전자상거래 플랫폼과 연계하여 사용자가 실시간으로 구매 가능한 형태로 제공한다. 사용자 경험의 즉시성과 상호작용성을 극대화하여 실시간 커머스 환경을 구현한다.

III. 결론

본 논문은 IPTV 실시간 방송 스트림 기반의 멀티모달 상품 정보 추론 및 스코어링 시스템을 제안하였다. 제안 시스템은 영상, 자막, 화면 내 텍스트의 통합 분석을 통해 방송 콘텐츠에 노출되는 상품을 실시간으로 인식하고, 신뢰도 점수 기반의 우선순위 추천정보를 제공하여 전자상거래 구매 시스템에 활용할 수 있도록 한다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광연구개발사업으로 수행되었음 (과제명: 전통예술 가무악의 융복합 공연제작 활성화를 위한 융복합 공연 기획/제작 플랫폼 기술 개발, 과제번호: RS-2024-00398536, 기여율: 100%)

참 고 문 헌

- [1] "Live video streaming and shopping interface integration system," US Patent US10491958B2, 2019.
- [2] "실시간 방송영상과 부가정보의 동기화 시스템," 한국특허 KR102123593B1, 2020.
- [3] "IPTV 기반 인터랙티브 영상 쇼핑 기술," 한국특허 KR101124560B1, 2012.