라이브 커머스 판매 예측을 위한 앵커-시청자 상호작용 기반 딥러닝 모델

정도현, 진승욱, 임지혜, 강금석*
한국과학기술원 데이터사이언스대학원, 한국과학기술원 경영공학부, 경희대학교,
*한국과학기술원 경영공학부

wjdehreh@kaist.ac.kr, tdns03@kaist.ac.kr, limjihye0323@gmail.com, *keumkang@kaist.ac.kr

Anchor-Viewer Interaction-Aware Deep Learning for Sales Prediction in Live Commerce

Dohyeon Jeong, Seungwook Jin, Jihye Lim, Keumseok Kang*
KAIST Graduate School of Data Science, KAIST College of Business, KyungHee Univ.,
*KAIST College of Business

요 약

본 논문은 라이브커머스 환경에서 실시간으로 발생하는 앵커와 시청자 간의 상호작용 데이터를 중심으로, 시계열적 흐름과 의미적 맥락을 반영하여 최종 판매액을 예측하는 딥러닝 모델을 제안한다. 기존 연구들은 주로 사전 수집된 정적 정보나 방송 중 생성되는 멀티모달, 광고, 시청자 행동 데이터를 활용한 예측에 초점을 맞추었다. 반면 본 연구는 실시간으로 관찰 가능한 상호작용 데이터를 시계열로 구성하고, 이를 LSTM과 양방향 Cross-Attention 구조를 통해 모델링하였다. 실험 결과, 제안된 모델은 기존 시계열 기반 예측 모델 대비 우수한 성능을 보였으며, 정적 정보와 결합한 실시간 상호작용 모델링으로도 유의미한 예측력을 확보함을 확인하였다. 이는 실시간 상호작용 정보를 활용한 판매 예측 가능성을 제시하고, 향후 메타버스와 같은 실시간 행위 중심 플랫폼으로의 확장 가능성을 보여준다.

Ⅰ. 서 론

라이브커머스는 실시간 스트리밍을 통해 제품을 소개하고 판매하는 형태의 전자상거래로, 정적 이미지와 텍스트 중심의 상품 설명, 후기 기반의 간접적 정보 전달 방식에 의존해왔던 기존 전자상거래와는 차별화된 소비자 경험을 제공한다. 라이브커머스에서는 앵커(판매자)가 실시간으로 상품을 시연하며 설명하고, 소비자는 실시간 채팅을 통해 질문하거나 반응을 표현하며 즉각적인 피드백을 받을 수 있다. 이와 같은 양방향 상호작용은 오프라인 쇼핑에 가까운 몰입감과 신뢰를 형성하며, 고객 만족도와 구매 전환율을 높이는 핵심요소로 작용한다.[1] 이러한 특징 덕분에 라이브커머스는 중국을 비롯한 아시아 지역에서 빠르게 확산되었으며, 국내에서도 다양한 플랫폼을 통해 여러 제품군에서 활발히 활용되고 있다.

기존 연구들에서는 이러한 라이브커머스 환경의 특성을 반영하여, 앵커, 상품, 방송 맥락, 시청자 데이터를 통합한 다양한 판매 예측 모델이 제안되어 왔다. 예를 들어, 앵커의 평판 정보와 상품 관련 텍스트·이미지 데이터를 멀티모달로 융합하여 판매액을 예측[2]하거나, 앵커, 상품, 방송 환경 정보를 각각의 엔티티 단위로 분리한 뒤 다양한 모달리티를 통합하는 방식으로 라이브커머스에서 발생하는 상호작용을 반영하여 예측 성능을 향상[3]시키고자 하였다. 또한, 일부 연구에서는 광고 데이터를 활용해 유입 트래픽을 먼저 예측한 후, 해당 트래픽과 방송 내 시청자의 행동(상호작용) 데이터를 단계별로 결합하는 Two-stage 구조를 통해 판매액을 예측하거나[4], 정량적인 데이터를 바탕으로 광고-방송-시청자 간의 상호작용을 고려하여 판매액을 예측하고자 하는 시도도 이루어졌다[5].

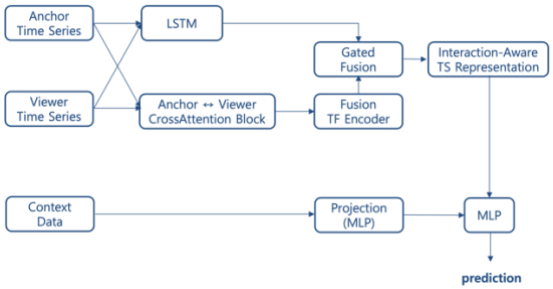
본 연구는 기존 연구들과 마찬가지로 라이브커머스 환경에서의 다양한 정보를 활용한 판매액 예측을 시도하지만, 기존 연구들이 앵커, 상품, 시청자 등 다양한 정보 간의 융합 및 상호작용에 초점을 맞추었다면, 본 연구는 특히 실시간 방송 중 상호작용의 중심이 되는 앵커와 시청자 간의 동태적 행위에 주목하고, 그 시간적 흐름과 맥락을 정밀하게 반영하고자 하였다. 기존에는 상호작용의 총량(예: 좋아요 수, 댓글 수, 주문 수 등)이나 정적인 특성에 주로 의존하였다면, 본 연구는 실시간 방송에서 이루어지는 앵커와 시청자의 행동 데이터를 시계열 형태로 구성하고, 이를 딥러닝 기반 예측 모델의 입력으로 활용하여 상호작용의 동적 양상을 모델 내부에서 구조적으로 반영한 판매액 예측 방식을 제안한다. 또한 광고 데이터나 상품 클릭 수, 주문 수와 같이 판매 성과와 직접적으로 연관되면서 수집 비용이 높은 정보 대신, API 및 스크립트 처리 등 자동화 가능한 방식으로 확보할 수 있는 실시간 상호작용 정보를 활용하여, 현실적인 제약 조건 하에서도 성능적으로 유의미한 예측 모델을 구축하고자 하였다.

Ⅱ. 본론

본 연구에서 활용한 데이터는 라이브커머스 방송 단위를 기준으로 구성되며, 크게 정적(context) 특성과 시계열(time-series) 특성으로 구분된다. 정적 특성은

방송 길이, 구성 상품 수, 푸시 알림 수, 라이브 할인 여부, 기획 방송 여부, 요일(주말 여부), 판매자 등급, 스토어 등급, 스토어 찜 수 등 각 방송의 고정된 속성을 포함하며, 판매액 예측에 필요한 맥락 정보를 제공한다. 시계열 특성은 방송 중 실시간으로 발생하는 앵커와 시청자의 행위를 1 분 단위로 수집한 데이터로, 상호작용의 흐름과 의미를 반영하기 위해 정량적 지표와 텍스트 임베딩이 함께 활용된다. 정량적 피처에는 앵커의 닉네임 호출 빈도, 발화속도, 앵커(매니저)의 댓글 수, 시청자의 댓글 수가 포함되며 이는 각 주체가 시도하는 상호작용의 강도나 빈도를 시간 축에 따라 정량화한 것이다. 텍스트 피처는 앵커의 발화와 시청자 댓글을 사전학습된 언어모델을 통해 임베딩으로 변환하여 포함하며[6], 수치 지표로 포착하기 어려운 정서적 분위기나 화제의 맥락을 보완적으로 반영한다. 이러한 정량적·의미 기반 피처의 결합을 통해 상호작용의 구조와 변화 양상을 보다 정밀하게 반영할 수 있도록 하였다.

한편 판매량, 할인 제공액, 방송 종료 시점의 상품 클릭 수, 좋아요 수 등과 같이 판매액과 직접적으로 연관되거나, 방송 종료 이후에야 확인 가능한 사후적 피처들은 제외하였다. 이러한 변수들은 예측 목표인 판매액과 높은 상관관계를 가져 예측 성능 향상에는 기여할 수 있으나, 실시간 상호작용 위주의 예측이라는 본 연구의 목적에는 적합하지 않으며, 실시간 예측 가능성과 실용성 확보를 위해 입력에서 배제하였다.



<그림 1. 모델 아키텍처>

예측 모델은 이렇게 구성된 정적 정보 및 시계열 정보를 통합하고, 실시간 상호작용의 동태적 특성을 효과적으로 반영하기 위해 LSTM 기반의 시계열 인코더와 Cross-Attention 기반의 상호작용 모델링, 그리고 Gated Fusion 및 정적 정보와의 결합 구조로 구성된다. 먼저, 앵커와 시청자의 시계열 데이터는 LSTM 에 입력되어 시간 흐름에 따른 행동 패턴을 인코딩하며, 두 시계열 데이터를 Cross-Attention 을 활용해 앵커와 시청자를 각각 Query, Key-Value 로 교차적으로 설정함으로써, 두 주체가 있음을 명시하고 상호작용의 방향성과 관계를 정교하게 학습할 수 있도록 하였다. Cross-Attention 의 출력을 Transformer Encoder 로 전달하여 전역 문맥과 장기 의존성을 반영한 표현을 정제하며, 이를 LSTM 기반 표현과 함께 Gated Fusion 모듈을 통해 통합한다. 해당 모듈은 학습된 게이트 값을 기반으로 두 표현 중 정보적으로 더 유의미한 부분에 가중치를 부여하여 최종 시계열 표현을 생성한다. 한편, 방송 단위의 정적 정보는 별도의 다층 퍼셉트론(MLP)으로 임베딩된 후, 시계열 표현과 결합되어 최종 예측에 활용된다. 이를 통해 모델은 실시간 상호작용의 시간적 흐름과 양방향성, 그리고 방송 맥락을 종합적으로 반영하여 라이브커머스 판매액을 예측할 수 있도록 설계되었다.

Ⅲ. 결론

본 연구는 라이브커머스 환경에서 실시간으로 발생하는 앵커와 시청자 간 상호작용 데이터를 중심으로, 이들의 시계열적 흐름과 의미적 맥락을 반영하여 최종 판매액을 예측하는 딥러닝 모델을 제안한다. 이를 위해 LSTM 기반의 시계열 인코딩, Cross-Attention 기반 상호작용 모델링, 정적 정보의 결합으로 구성된 구조를 설계하였으며, 상호작용의 시간적 특성과 방향성을 효과적으로 포착할 수 있게 하였다.

Prediction model	Evaluation metrics		
	RMSE	MAE	R^2
RNN	0.5203 ± 0.0095	0.3999 ± 0.0080	0.7241 ± 0.0102
LSTM	0.5171 ± 0.0104	0.3927 ± 0.0093	0.7274 ± 0.0110
GRU	0.5123 ± 0.0167	0.3930 ± 0.0119	0.7323 ± 0.0179
Transformer	0.4972 ± 0.0127	0.3841 ± 0.0117	0.7479 ± 0.0130
LSTM+Transformer Fusion	0.4932 ± 0.0093	0.3811 ± 0.0080	0.7520 ± 0.0093
LSTM+Cross-Attention Fusion	0.4710 ± 0.0063	0.3622 ± 0.0035	0.7738 ± 0.0060

<표 1. 예측 모델별 성능 비교 (RMSE, MAE, R^2 기준)>

실험 결과, 제안된 모델은 기존 시계열 예측 모델(RNN, LSTM, GRU, Transformer, LSTM+Transformer Fusion) 대비 전반적으로 우수한 예측 성능을 보였으며, 실시간 상호작용 정보를 Cross-Attention 구조를 통해 정교하게 반영함으로써 예측 성능 향상에 기여하였다. 또한 광고 및 구매 관련 시청자 행동과 같은 수집 비용이 높은 외부 정보를 활용하지 않고도, 방송 중 실시간으로 외부에서 관찰 가능한 시청자 댓글과 앵커 발화 내용을 기반으로 상호작용을 구성하고 이를 사전 확보 가능한 정적 정보와 결합하여 유의미한 예측성능을 달성하였다. 이러한 결과는 실시간 상호작용 정보의 구조적 활용을 통해 효과적인 판매 예측이 가능함을 보여주며, 향후 메타버스나 가상 환경 등에서 인간·객체 간 실시간 상호작용이 중심이 되는 다양한 플랫폼으로의 예측 모델 확장 가능성을 시사한다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학 ICT 연구센터(ITRC)의 지원을 받아 수행된 연구임(IITP-2025-RS-2021-II211816)

참 고 문 헌

[1] Yun, Jeewoo, et al. "New generation commerce: The rise of live commerce (L-commerce)." Journal of Retailing and Consumer Services 74 (2023): 103394.

[2] Xu, Wei, Ying Cao, and Runyu Chen. "A multimodal analytics framework for product sales prediction with the reputation of anchors in live streaming e-commerce." Decision Support Systems 177 (2024): 114104.

[3] Xu, Guang, et al. "MEMF: Multi-entity multimodal fusion framework for sales prediction in live streaming commerce." Decision Support Systems 184 (2024): 114277.

[4] Lin, Qinping, et al. "A two-stage prediction model based on behavior mining in livestream e-commerce." Decision Support Systems 174 (2023): 114013.

[5] Wang, Lijun, and Xian Zhang. "Livestream sales prediction based on an interpretable deep-learning model." Scientific Reports 14.1 (2024): 20594.

[6] Sturua, Saba, et al. "jina-embeddings-v3: Multilingual embeddings with task lora." arXiv preprint arXiv:2409.10173 (2024).