개인화 가전의 사용자 맞춤형 서비스 제공을 위한 온디바이스 AI 기반 사용자 행동 예측 모델 개발

김소현, 최원기, 정영환, 이상신*

한국전자기술연구원

{shk980903, cwk1412, cjstntjd}@keti.re.kr, *sslee@keti.re.kr

Implementation of an On-Device AI-Based User Behavior Prediction Model for Personalized Home Appliance Services

So Hyeon Kim, Wongi Choi, Younghwan Jeong, Sangshin Lee*

Korea Electronics Technology Institute

요 약

최근 가전 제품은 사용자의 다양한 생활 패턴을 반영하여 맞춤형 서비스를 제공하는 지능형 시스템으로 발전하고 있다. 이러한 개인화 가전은 저전력 환경에서 사용자의 선호 및 행동을 예측하여 능동적으로 제어하는 기술이 중요하며, 동시에 개인정보 보호 및 하드웨어 특성을 고려한 온디바이스 AI의 적용이 요구되나 연산 자원이 제한된 온디바이스 환경에서 복수의 제어 요소를 동시에 예측하는 데에는 한계가 있다. 본 논문에서는 개인화 가전의 사용자 맞춤형 서비스 제공을 위해, AR Transformer 모델을 활용한 시계열 사용자 행동 예측 시스템을 구현한다. 사용자 이용 로그를 기반으로 모델을 훈련하고 온라인 업데이트된 단일 모델을 활용하여 사용자 선호 행동을 다중 예측한다. 실험 결과 제안하는 방법은 저전력 디바이스 환경에서도 기존 시계열 예측 모델 대비 복잡한 사용자 활용 패턴을 상대적으로 잘 포착하는 것으로 확인되었다.

Ⅰ. Introduction

인공지능과 사물인터넷(IoT)의 확산으로 가전제품은 사용자의 상태와 선호를 이해하고 능동적으로 반응하는 지능형 시스템으로 발전하고 있다. 특히 안마의자와 같은 개인화 가전은 사용자별 신체 상태, 피로도, 시간대에 따라 사용 패턴과 선호 모드가 크게 달라지기 때문에, 동일한 기기라 하더라도 개별 사용자에게 최적화된 동작을 제공하는 기술의 중요성이 높아지고 있다. 기존에는 클라우드 기반 AI가 대규모 로그 데이터를 수집·분석하여 자동 모드나 추천 설정을 제공하는 방식이 주류를 이루었다. 그러나 네트워크 지연, 연결 불안정, 개인정보 전송에 따른 보안 문제 등으로 인해, 실시간 반응성과 안정성이 요구되는 가전 환경에서는 클라우드 중심 구조만으로 충분한 품질을 보장하기 어렵다. 이러한 한계로 인해 모델 추론을 기기 내부에서 직접 수행하는 온디바이스 AI(on-device AI)가 중요한 대안으로 부상하고 있으며, 개인화 가전 분야에서도 핵심 기술로 자리 잡고 있다.[1]

한편, 가전 사용 로그는 다양한 제어 요소가 시간 흐름에 따라 변화하는 시계열 데이터로 구성되며, 실제 사용 과정에서는 모드가 언제 전환되는지(전환 시점)와 어떤 설정으로 전환되는지(전환 결과)가 사용자 경험을 결정하는 중요한 요인이 된다. 그러나 단순한 시계열 모델은 이러한 전환 구간을 충분히 포착하지 못하거나, 이전 상태를 그대로 복사하는 방향으로 모델이 포화되는 문제가 흔히 발생한다. 더불어 온디바이스 환경의 제한된 연산·메모리 자원은 복잡한 대규모 모델의 적용을 어렵게 하므로, 경량이면서도 전환 패턴을 명확히 학습할 수 있는 맞춤형 구조가 필요하다.

이에 본 논문에서는 개인화 가전의 사용자 맞춤형 서비스 제공을 위해, 온디바이스 환경에 적합한 자동회귀(autoregressive) Transformer 기반 사용자 행동 예측 모델을 제안한다.[2] 제안 모델은 사용자 ID, 시간대, 시작 시각과 같은 메타데이터를 컨텍스트로 활용하여, 정규화된 향후 일정 구간의 제어 설정을 순차적으로 예측한다. 특히 모드 전환 여부와 전환 결과를 분리하여 학습하는 구조를 도입함으로써, 실제 사용 과정에서 중요한 전환 구간을 명시적으로 모델링하고, 단순 복사 경향을 완화하도록 설계하였다. 또한 전환 시점에 더 높은 학습 가중치를 부여하는 방식으로 희귀한 전환 이벤트를 효과적으로 학습하며, 전체 모델은 온디바이스 실행을 고려한 경량 Transformer 구조로 구성하여 실시간 추론 가능성과 자원 효율성을 확보하였다.

본 연구는 이러한 접근을 통해 사용자별·상황별 행동 패턴을 반영하는 동시에, 자원 제약이 존재하는 가전 기기에서 실시간으로 동작 가능한 사용자 행동 예측 모델의 설계 가능성과 실용성을 검증하는 데 목적이 있다. 이를 통해 개인화 가전 분야에서 요구되는 맞춤형 서비스, 프라이버시 보호, 온디바이스 AI 실행 가능성의 세 가지 요구를 동시에 충족하는 모델 설계 방향을 제시한다.

Ⅱ. Proposed Method

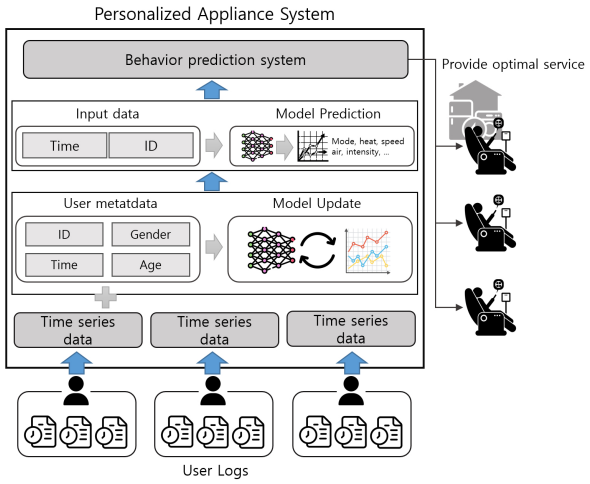
본 연구는 개인화 가전의 사용자 행동을 온디바이스 환경에서 실시간으로 예측하기 위해, 전환 중심 구조(mode-change aware)로 설계된 경량 Autoregressive Transformer 모델을 제안한다. 제안하는 전체 시스템의 개념적 구조는 그림 1에 나타나 있으며, 사용자 로그로부터 시계열 데이터를 구성하고, 사용자 메타데이터와 결합하여 행동 예측 모델을 업데이트한 뒤, 최종적으로 온디바이스에서 개인별 최적 서비스를 제공하는 흐름으로 이루어진다.

우선, 실제 사용 로그는 이벤트 간의 시간 간격이 일정하지 않기 때문에, 각 세션을 고정 길이 시퀀스로 변환하는 정규화 과정을 수행한다. 사용자가 특정 모드나 설정에 머무른 시간 비율을 반영하는 time-weighted 방식으로 대표 상태를 결정하여, 모드·강도·공기압·온열로 구성된 균일 길이의 시계열을 생성한다. 추가적으로 사용자 ID, 시간대(morning/lunch/evening), 시작 시각을 메타 정보로 활용하며, 시작 시각은 하루 단위의 주기성을 반영할 수 있도록 Fourier 기반의 특징으로 변환하여 모델 입력에 포함한다.[3]

제안 모델은 이와 같이 구성된 입력을 통합하여 하나의 컨텍스트 벡터로 변환하고, Transformer Encoder를 통해 세션의 전반적 특징을 단일 메모리 형태로 요약한다. 이후 Transformer Decoder는 teacher forcing 기반의 자동회귀(autoregressive) 방식을 사용하여, 이전 시점의 실제 상태를 입력으로 받아 향후 여러 스텝의 설정 패턴을 순차적으로 생성한다. 디코더 출력은 모드, 강도, 공기압, 온열 설정을 각각 예측하는 head와, 해당 시점이 모드 전환이 발생하는지 여부를 예측하는 이진 분류 head로 구성된다. 이처럼 전환 시점과 전환 결과를 분리한 다중 head 구조는 실제 사용 과정에서 중요한 전환 구간을 명확하게 학습하게 하며, 전통적 시계열 모델이 갖는 단순 복사 경향을 줄이는 데 기여한다.

전환 구간을 보다 효과적으로 학습하기 위해 손실 함수 또한 전환 중심으로 설계하였다. 모드 예측에는 전환이 발생하는 시점에 더 큰 가중치를 부여하고, 부수 제어 요소(intensity·air·heat)는 전환 순간의 가중치를 낮추어 학습 간섭을 줄였다. 모드 전환 이벤트가 상대적으로 적어 발생하는 불균형 문제는 양성 클래스 가중치(pos-weight) 혹은 focal 방식의 조정을 통해 완화함으로써, 희귀 전환 패턴을 안정적으로 학습하도록 하였다.

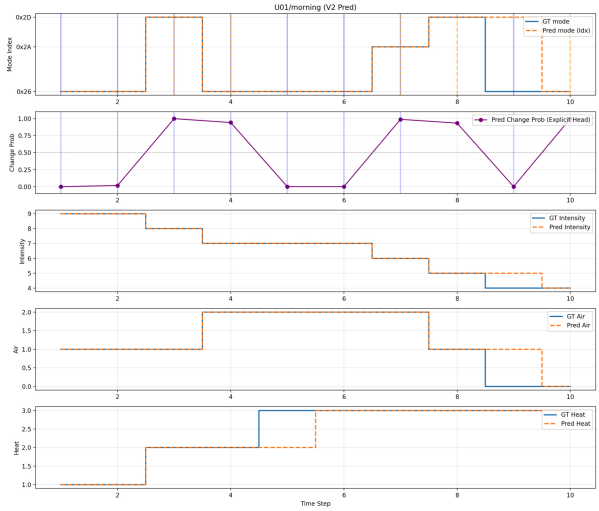
전체 모델은 온디바이스 환경에서의 실시간 추론을 고려하여 경량 Transformer 구조를 채택한다. 컨텍스트 정보를 단일 토큰으로 구성하여 Encoder 연산량을 최소화하고, 모델 차원 및 레이어 수·헤드 수를 제한하여 연산 자원을 절감하였다. 이와 같은 설계를 통해 제안 모델은 사용자별·시간대별 행동 패턴을 반영하면서도 제한된 계산 능력을 가진 가전 기기에서도 안정적으로 동작할 수 있는 효율성을 확보한다.



(그림 1) 사용자 행동 예측 시스템 구조도

Ⅲ. Simulation Result

제안한 전환 중심 Autoregressive Transformer 모델의 예측 성능을 확인하기 위해, 실제 사용자 로그를 기반으로 생성된 정규화 세션에 대해 시뮬레이션을 수행하였다. 그림 2는 사용자 활용 로그에 대해 모델이 예측한 모드, 강도, 공기압, 온열 설정의 시계열 결과를 GT(Ground Truth)와 비교하여 나타낸 것이다. 각 그래프는 시간 흐름에 따른 제어 상태 변화를 시각적으로 비교할 수 있도록 구성하였다. 그림에서 확인할 수 있듯이, 제안 모델은 전반적인 시퀀스 흐름뿐 아니라 전환 구간을 비교적 정확하게 포착하고 있다. 특히 모드(mode)의 경우, 사용자가 실제로 설정을 변경한 시점에서 예측 결과가 전환 패턴을 따라가며, 전환 직후의 모드 유지 구간에서도 안정적인 예측을 보여준다. 이는 본 연구에서 도입한 전환 시점 분리 예측(head separation)전환 중심 가중 손실 설계가 효과적으로 작동했음을 의미한다. 강도(intensity)와 공기압(air) 설정 또한 GT와 거의 일치하는 추세를 보이며, 단계적 증가·감소 패턴, 유지 구간 등이 정확하게 예측되었다. 특히 강도의 경우, 단계적으로 감소하는 추세와 유지 구간이 전반적으로 잘 반영되었으며 전체적인 사용 패턴을 안정적으로 재현하고 있다. 온열(heat)은 세션 전반에 걸쳐 변화 구간이 거의 없는 특성을 보였고, 예측 결과 또한 실제 활용 패턴과 유사하게 불필요한 전환 없이 유지되는 경향을 나타낸다. 이는 실제 사용자 조작 빈도와 패턴을 반영하여 예측을 수행하고 있음을 보여준다.



(그림 2) 사용자 행동 예측 결과 비교 그래프

Ⅲ. Conclusion

본 연구에서는 개인화 가전의 사용자 행동을 온디바이스 환경에서 실시간으로 예측하기 위한 경량 Autoregressive Transformer 모델을 제안하였다. 제안 모델은 실제 사용 로그의 불규칙한 시간 간격을 정규화하고, 사용자 메타데이터와 시간 정보를 통합하여 시퀀스 기반 예측이 가능하도록 설계되었다. 특히 모드 전환 여부와 전환 결과를 분리하여 학습하는 구조와 전환 중심 가중 손실 전략을 도입함으로써, 사용 과정에서 중요한 전환 구간을 명확하게 포착할 수 있도록 하였다. 시뮬레이션 결과, 제안 모델은 모드·강도·공기압·온열과 같은 다양한 제어 요소를 높은 정밀도로 예측하였으며, 전환 구간에서도 실제 사용자 패턴과 거의 일치하는 결과를 보였다. 이는 전환 중심 구조와 경량 Transformer 설계가 효과적으로 결합되어, 온디바이스 환경에서도 충분한 예측 성능을 확보할 수 있음을 의미한다. 추후 연구에서는 더 다양한 사용자군에 대한 일반화 성능을 검증하고, 장기 예측 모델링 및 지능형 제어 시스템과의 통합을 포함한 후속 연구를 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 2026년도 정부(산업통상자원부)의 재원으로 한국산업기술기획평가원의 지원을 받아 수행된 연구임 (RS-2024-00507555, 국산 SoC 기반 온디바이스 AI 초개인화 가전 시스템 개발 실증, 기여율 100%)

참 고 문 헌

[1] Wang, X., “Empowering Edge Intelligence: A Comprehensive Survey on On-Device AI Models”, Art. no. arXiv:2503.06027, 2025.

[2] S. Mohammad, G. Song, and T. Zhu, “EdgeFlex-Transformer: Transformer Inference for Edge Devices,” arXiv:2512.19741, 2025.

[3] Li, Wenxiang & Law, K.. (2024). Deep Learning Models for Time Series Forecasting: A Review. IEEE Access. PP. 1-1. 10.1109/ACCESS.2024.3422528.