고정형 헬퍼와 추천 시스템을 활용한 강화학습 기반 인센티브 매커니즘 연구

김대현, 김기현, 임민중
동국대학교

eoguseoeo@naver.com, rlsrlgus123@naver.com, minjoong@dongguk.edu

A Study on Reinforcement Learning-Based Incentive Mechanism Using Fixed Helper and Recommendation Systems

Dae-hyun Kim, Ki-Hyoen Kim, Minjoong Rim
Dongguk University

요 약

본 논문은 무선 네트워크의 커버리지와 용량 문제를 해결하고 사용자 경험을 향상시키기 위해 D2D (Device-to-Device) 캐싱 시스템의 효과적인 구현에 초점을 맞추고 있다. 기존 연구에서는 인센티브 제공이 사용자의 캐싱 행위를 촉진하고 네트워크 효율성을 개선하는데 효과적임을 입증하였다. 하지만, 단순한 인센티브 제공 이상의 전략이 필요함을 인식하고, 본 연구에서는 인센티브를 보다 효율적으로 분배하는 방안을 제안한다. 인센티브 분배에 초점을 맞춘 새로운 접근 방식을 통해 D2D 캐싱 시스템과 추천 시스템의 성능을 향상시키고자 한다. 주요 목표는 사용자들의 참여를 유도하고 캐시 적중률을 증대시키는 것으로, 인공지능 강화학습을 기반으로 한 인센티브 분배 방안을 연구한다.

Ⅰ. 서 론

현대 사회에서 정보와 통신 기술은 끊임없이 진화하고 있으며, 이동통신 시스템은 그 중심에 서 있다. 특히, 비디오 스트리밍 서비스의 급증과 데이터 사용량의 증가는 이동통신 시스템에 상당한 부담을 주고 있으며, 이로 인해 네트워크 과부하 문제가 심각한 관심사로 부상하고 있다. 이러한 문제를 해결하기 위해, 과거의 연구들은 네트워크 대역폭의 효율적 관리, 데이터 전송의 최적화, 콘텐츠 캐싱과 같은 다양한 접근 방식을 제시하였다. 특히, 캐싱 네트워크 구성 방법은 네트워크 성능을 향상시키고 사용자의 콘텐츠 접근성을 높이는 데 중요한 역할을 하였다. [1]

사용자의 콘텐츠 요구와 행동 패턴을 분석하고 이를 기반으로 최적의 콘텐츠를 제공하는 추천시스템은 디지털 플랫폼에서 핵심적인 역할을 수행하고 있다. 추천시스템의 발전은 사용자 경험을 향상시키는 동시에, 캐싱 시스템의 효율성을 높이는 데에도 기여하고 있다. 더 나아가, 인센티브 기반 캐싱 시스템은 사용자들의 참여를 촉진하고 캐시 적중률을 향상시키는 새로운 접근 방식으로 주목받고 있다. 이러한 시스템은 사용자들에게 캐싱에 참여하도록 동기를 부여함으로써, 네트워크의 부담을 줄이고 전체적인 성능을 개선하는 효과를 가져온다. D2D 캐싱 및 추천 시스템의 통합은 사용자 데이터의 효율적인 활용과 네트워크 자원의 최적화를 가능하게 함으로써, 이동통신 시스템의 용량 및 성능을 대폭 향상시킬 수 있다. [2]

본 연구는 D2D 캐싱 시스템과 추천 시스템의 상호작용에 초점을 맞추고 있다. D2D 캐싱 시스템은 사용자 간의 직접적인 데이터 공유를 가능하게 하여 콘텐츠 접근성을 높이며, 추천 시스템은 사용자의 취향과 관심사를 분석하여 적합한 콘텐츠를 추천함으로써 캐싱 및 데이터 공유를 촉진한다. 이러한 상호 작용은 네트워크 성능의 향상뿐만 아니라, 사용자 경험의 개선에도 기여할 것으로 기대된다. 본 연구는 이동통신 시스템의 과부하 문제 해결, 캐싱 네트워크 구성 방법의 최적화, 그리고 추천 시스템의 효율성 개선에 기여하는 새로운 접근 방식을 제안하며, 이를 통해 네트워크의 효율성 및 사용자의 만족도를 향상시키는 방안에 대해 논의하고자 한다.[3]

Ⅱ. 본론

본 논문에서 제안하는 모바일 데이터 트래픽의 병목 현상 및 서비스 지연 문제 해결 방안은 디바이스 간 직접 통신(D2D 통신)과 대용량 캐시 시스템인 헬퍼를 활용한 캐싱 시스템 설계에 중점을 둔다. 강화학습 기반의 인센티브 분배 모델을 통해 캐시 시스템의 효율성을 극대화하고자 한다. 이 절에서는 강화학습을 활용한 인센티브 분배 모델에 대해 설명하며, 이를 통한 캐시 시스템의 성능 향상 방안을 제시한다.

강화학습은 에이전트가 환경과 상호작용하며 보상을 최대화하는 방향으로 행동을 학습하는 기계학습의 한 분야이다. 에이전트는 시행착오를 통해 어떤 상태에서 어떤 행동을 취할 때 가장 많은 보상을 받을 수 있는지를 학습한다. 본 논문에서는 인센티브 분배 모델의 구현을 위해, 본 논문에서는 강화학습 알고리즘을 적용하여 시스템이 최적의 인센티브 분배 전략을 학습할 수 있도록 한다. 강화학습에서 에이전트는 주어진

상태(S)에서 어떤 행동(A)을 취했을 때 얻을 수 있는 보상(R)을 기반으로 최적의 행동을 결정한다. 이 과정에서, 에이전트는 시행착오를 통해 여러 상황에서의 최적의 행동을 탐색하고 학습하게 된다. 상태(S): 시스템의 현재 상태로, 사용자의 콘텐츠 선호도, 캐시의 현재 상태, 네트워크 상황 등을 포함한다. 행동(A): 인센티브를 어떻게 분배할지에 대한 결정이다. 예를 들어, 특정 헬퍼나 UE에 인센티브를 얼마나 줄 것인지 등이다. 보상(R): 행동의 결과로 받는 보상으로, 캐시 적중률의 향상, 데이터 트래픽의 감소, 사용자 만족도 증가 등을 기준으로 측정된다.

본 연구에서는 사용자의 선호도와 인센티브에 대한 반응 민감도를 모델링하기 위해 강화학습 방법 중 하나인 DDPG(Deep Deterministic Policy Gradient) 알고리즘을 사용하였다. DDPG 는 연속적인 행동 공간에서 효과적으로 작동하는 알고리즘으로, 정책 기반과 가치 기반 강화학습의 장점을 결합한 것이 특징이다. DDPG 알고리즘의 적용은 캐시 시스템 내에서 헬퍼와 사용자 장비(UE)에게 동적으로 인센티브를 분배하는 과정에서 중요한 역할을 한다. 본 연구는 시스템의 현재 상태를 기반으로 최적의 인센티브 분배 전략을 결정할 수 있다. 이는 다음과 같은 과정을 통해 수행된다. 첫째, 알고리즘은 현재 시스템 상태를 입력으로 받아, 각 헬퍼와 UE 에게 얼마만큼의 인센티브를 할당할지 결정하는 정책을 출력한다. 이때, DDPG 의 액터(정책) 네트워크가 이 역할을 수행한다. 둘째, 결정된 인센티브 분배 전략의 성과를 평가하기 위해, 시스템은 해당 전략을 실행하고 그 결과를 관찰한다. 이 결과는 캐시 적중률의 향상, 데이터 트래픽의 감소, 사용자 만족도 증가 등으로 측정되며, 이를 통해 보상을 계산한다. 셋째, 이 보상을 기반으로 DDPG 의 크리틱(가치) 네트워크는 액터 네트워크가 결정한 정책의 가치를 평가하고, 이를 통해 액터 네트워크의 성능을 개선한다. 이러한 과정을 반복하면서, DDPG 알고리즘은 점차 최적의 인센티브 분배 전략을 학습해 나간다. 기존의 방법들은 주로 사용자의 선호도만을 고려한 반면, 본 연구에서는 인센티브에 대한 반응 민감도를 포함하는 복잡한 사용자 선호도 벡터를 소개하고자 한다. 이를 위해 softmax 함수를 기반으로 한 변형된 모델을 개발하였다. 제안하는 모델은 사용자 u 의 선호도 벡터를 $z_u = z_{u1}, \ldots, z_{uM}$ 로 나타내며, 여기서 M 은 인센티브의 비율을 의미한다. 또한, 각 인센티브 i 에 대한 사용자 u 의 반응 민감도를 포함하는 인센티브 반응 벡터를 $r_u = r_{\{u1\}}, \ldots, r_{\{uM\}}$ 로 정의한다. 사용자 u 의 인센티브 i 에 대한 선호도 점수 q_{ui} 는 다음과 같이 계산된다.

$$q_{ui} = \frac{\exp(z_{ui} + r_{ui})}{\sum_{j=1}^{M} \exp(z_{uj} + r_{uj})} \quad (1)$$

z_{ui} 는 사용자 u 의 인센티브 i 에 대한 기본 선호도를, r_{ui} 는 해당 인센티브에 대한 사용자의 반응 민감도를 각각 나타낸다. 분모는 모든 가능한 인센티브에 대하여 $z_{uj} + r_{uj}$ 의 지수 함수의 합을 계산함으로써, q_{ui} 가 사용자 u 의 모든 인센티브에 대한 선호도를 정규화하도록 한다. 이러한 접근 방식을 통해, 본 연구는 사용자의 선호도뿐만 아니라 특정 인센티브에 대한 반응 민감도까지 고려하여 각 사용자에 대한 선호도 벡터를 보다 세밀하게 모델링할 수 있다. 이를 통해 계산된 q_{ui}는 사용자 u가 인센티브 i를 선호할 확률을 나타낸다.

강화학습 모델을 사용하여 캐시 시스템 내에서 헬퍼와 UE 에게 인센티브를 동적으로 분배한다. 인센티브의 총량은 시스템의 정책에 따라 사전에 결정되며, 각각의 헬퍼와 UE 에게 할당되는 인센티브의 양은 시스템의 상태와 환경 변화에 따라 강화학습 알고리즘을 통해 조절된다. 이를 통해 헬퍼는 캐시 성능을 극대화하고, UE 는 효율적인 콘텐츠 소비를 통해 데이터 사용량을 줄일 수 있다. 강화학습을 적용한 실험에서는, 추천 항목의 수 K 가 500 개이고, 캐시 크기N_cache가 2000, 그리고 사용자가 실제로 요청할 확률 α가 0.6 일 때의 조건을 기반으로 캐싱 및 추천 시스템의 성능 지표 B 값과 캐시 적중률을 분석했다.

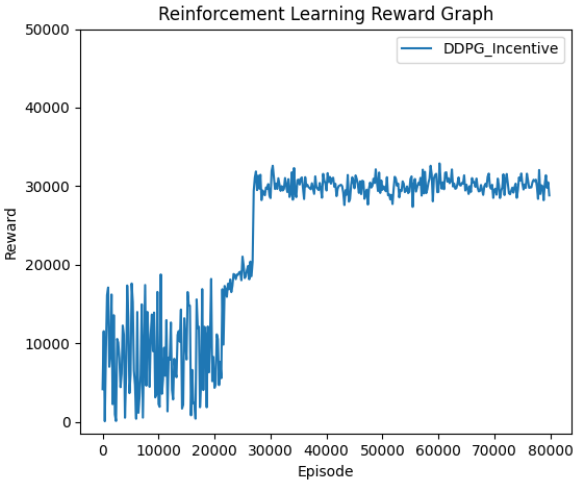

그림 1. 강화학습 DDPG 알고리즘 기반 인센티브 분배를 통한 사용자의 선호도 학습 모델

Ⅲ. 결론

강화학습 기반 인센티브 분배 모델은 캐시 시스템의 성능을 극대화할 수 있으나, 알고리즘의 복잡성과 한정된 상태 고려로 인해 현실적 모델링에 어려움이 있다. 이 모델은 시스템의 적응성과 사용자 만족도를 높이며 모바일 데이터 트래픽 문제를 해결할 수 있지만, 실제 적용을 위해서는 알고리즘 복잡도 해결과 실제 상황 반영의 개선이 필요하다. 따라서, 향후 연구에서는 이러한 한계를 극복하기 위한 방안 모색이 중요할 것이다.

ACKNOWLEDGMENT

본 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2022R1F1A1062987).

참 고 문 헌

[1] 김대현, 김기현, 임민중, D2D 캐싱 시스템에서 사용자 선호도를 고려한 인센티브 분배 기법, 한국통신학회 동계학술대회 2024.

[2] Dongsheng Zheng, Yingyang Chen et al., “Cooperative Cache-Aware Recommendation System for Multiple Internet Content Providers” IEEE Wireless communications Letters, Vol. 9, No. 12, December 2020.

[3] L. Li, G. Zhao, and R.S. Blum, "A Survey of Caching Techniques in Cellular Networks: Research Issues and Challenges in Content Placement and Delivery Strategies," IEEE Commun. Surveys & Tutorials, vol.20, no.3, pp.1710-1732, Third Quarter 2018.