

콘텐츠 캐싱을 위한 기계학습 기반 이동성 예측 기법 연구

임한여름⁰, 홍충선*

경희대학교 컴퓨터공학과

{prosummer, cshong}@khu.ac.kr

A Research on Machine Learning Based Mobility Prediction Method for Content Caching

HanYeoReum Im⁰, ChoongSeon Hong*

Department of Computer Science and Engineering, Kyung Hee University

요 약

최근 방대한 모바일 트래픽의 지연을 방지하고자 사용자와 가까운 위치에서 연결 및 송수신을 담당하는 Small-cell Base Station에 콘텐츠를 캐싱하는 방안이 대두되고 있다. 반면, 콘텐츠 중심 네트워크에서 콘텐츠 캐싱의 효율성을 높이기 위해 사용자의 이동성을 고려하는 새로운 움직임이 커지고 있다. 따라서 본 논문은 기계학습 기법인 LSTM을 활용해 이동성을 예측하는 방안을 제시해 향후 효율적인 콘텐츠 캐싱 및 자원 할당이 가능하도록 하였다.

1. 서 론

최근 다양한 모바일 기기와 IoT의 부상으로 모바일 트래픽 양이 급증하면서 대역폭에 대한 방대한 수요를 네트워크 밀집화를 통해 충족하고자 하는 시도가 이루어지고 있다. 반면, 5G와 같이 대역폭이 증가하게 되는 경우 백홀망(backhaul link)의 혼잡도가 증가하여 모바일 트래픽이 지연되는 결과를 가져올 수 있다. 따라서 콘텐츠 중심 통신 트렌드에 발맞춰 SBSs(Small Cell Base stations)에 콘텐츠를 미리 캐싱해둠으로써 백홀망 혼잡도를 낮추는 방안이 대두되고 있다 [1]. 최근 SSDs(Solid State Drives)와 같은 캐싱 디바이스의 가격이 지속적으로 저렴해지면서 비용적인 측면에서도 백홀망보다 사용자측에 가까운 SBSs에 캐싱하는 것이 보다 유리하다 [2].

CCWNs(Content-Centric Wireless Networks)와 같은 콘텐츠 중심 네트워크에서 중요하게 다뤄질 수 있는 feature로 사용자 이동성을 고려할 수 있다. 사용자의 이동성을 미리 알 수 있다면 이를 활용한 이동성 관리(mobility management)를 넘어 자원 할당에도 효과적으로 활용할 수 있다. 따라서 이동성을 예측하여 콘텐츠를

캐싱하면 백홀망을 거쳐 코어 네트워크까지 데이터를 송수신하지 않고 사용자와 가까운 SBSs에 캐싱된 데이터를 바로 받을 수 있으므로 service latency를 눈에 띄게 낮출 수 있다. 이에 본 논문은 콘텐츠 캐싱의 Cache-Hit Ratio를 높여 궁극적으로 Energy efficiency를 높이고자 기계학습 기법을 기반으로 사용자의 이동성을 예측하는 것을 제안한다.

2. 관련 연구

◎Markov Renewal Process

Markov renewal process란 Markov jump process 개념을 일반화한 random process로 시간에 따른 상태 공간에서의 random variables 간의 시간차에 해당하는 jump time을 고려한 통계적 시행이다. [3]에서는 사용자 이동성 예측과 관련된 연구로 사용자의 이동 경로와 체류 시간을 고려하여 Markov renewal process를 모델링한 것을 Base station뿐만 아니라 D2D(Device-to-Device) 환경까지 적용하였다.

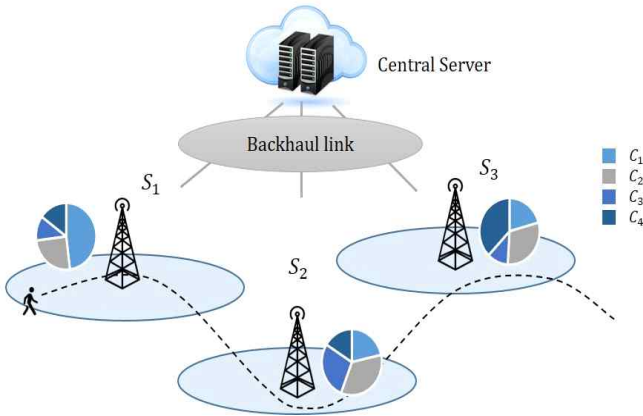
◎Machine Learning

통계학적 방법뿐 아니라 기계학습을 활용한 사례로 RNNs(Recurrent Neural Networks)보다 장기적인 예측을 가능하게 하는 LSTM(Long Short Term Memory)을 활용한 연구가 있다. 해당 연구[4]에서는 사용자 이동성을 예측하고 일정한 시간 간격 내에 사용자 군집을 형성해 sub-channel을 최적으로 할당해 network throughput을 눈에 띄게 향상시켰다.

이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017R1A2A2A05000995). 아울러 본 연구는 한국정보화진흥원(NIA)의 미래네트워크선도시험망(KOREN) 사업 지원과제의 연구결과로 수행되었음 (18-951-00-001). *Dr. CS Hong is the corresponding author.

3. 제안 사항

3.1 시스템 모델 및 시나리오



[그림 1] 사용자의 이동성을 고려한 SBSs에서의 협력 캐싱

본 연구의 시스템 모델은 [그림 1]과 같다. Central Server와 backhaul link를 통해 연결된 SBSs가 S1, S2, S3 와 같이 있다고 가정하자. 그리고 각 SBSs의 range를 이동하는 사용자 U1을 가정했을 때, 시나리오는 다음과 같다.

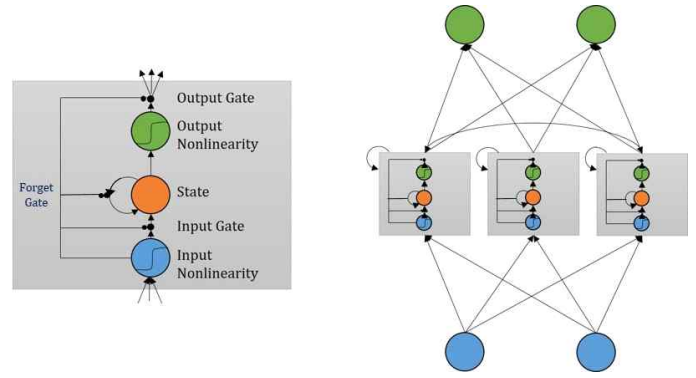
- (1) 사용자 이동 패턴 예측
- (2) 예측값 기반 사용자 군집
- (3) 각 군집 별 맞춤 콘텐츠 배치

위의 시스템 모델을 예로 설명하면, U1과 같은 사용자들의 이동성 및 Base station 사용 데이터를 활용하여 사용자의 이동성 즉, 이동 패턴을 LSTM을 활용하여 예측한다. 이동 패턴의 예측은 Central Server에서 이루어지는 것으로 가정한다. 여기서 예측된 각각의 이동 패턴 중 비슷한 패턴끼리 군집을 형성하고 이 군집을 활용하여 각 군집의 특징에 맞게 콘텐츠를 캐싱하는 것이다. 이러한 시나리오 안에서 본 논문은 사용자 이동 패턴 예측 과정을 중점적으로 다루는 것을 목적으로 한다.

3.2 기계학습 기반 이동성 예측

3.2.1 LSTM

시계열 데이터를 예측할 때 기계학습을 활용하면 빠르고 정확한 결과를 얻을 수 있다. Hochreiter와 Schmidhuber에 의해 제안된 LSTM은 기울기가 사라지거나 폭발하는 문제를 극복하는 특별한 RNN 구조로 장기간의 의존 관계를 학습하는데 탁월한 능력이 있다. 즉 오랜 기간에 걸친 데이터 로그를 분석할 때 효과적이다. LSTM 셀은 내부적으로 여러 단계에 걸친 에러를 기억하는데 내부 상태가 1로 고정된 웨이트와 선형 활성화 함수와 연결되어 있기 때문에 지역 미분값은 항상 1이 된다는 특징이 있다[5].



[그림 2] Long-Short Term Memory 구조

3.2.2 알고리즘

Algorithm 1. User Mobility Prediction Model

Input: Training Dataset d_1 , Test Dataset d_2

Output: Trained model m^* , Predicted value of Test data

▷ Model Training

- 1: Initialize learning rate l_m , sequence_length s_m , the number of LSTM cells α_m , the number of LSTM layers β_m
- 2: load Training dataset d_1 , Test dataset d_2
- 3: **for** data in Training dataset d_1 and Test dataset d_2 **do**
- 4: generate and store sequence data
- 5: build LSTM model m with α_m, β_m
- 6: **while** training accuracy is not large enough **do**
- 7: minimize the training loss of model m with cross-entropy loss function and Adam optimizer
- 8: **if** validation accuracy of model m is better than previous model **then**
- 9: update model m

▷ Model Testing

- 10: **for** sequence data of Test dataset d_2 **do**
- 11: predict next values with model m^*

Algorithm 1은 central server에서 수행되는 예측 모델의 pseudo code를 나타낸다. 해당 모델은 학습을 위한 Dataset d_1 를 LSTM 기반으로 학습시켜 최적의 모델 m^*

을 얻고 그 모델로 테스트 데이터를 정확하게 예측하는 것을 목적으로 한다. (line 1)에서 학습률, 시퀀스 길이, LSTM cell과 layer의 개수를 초기화한다. (line 3 - line 4)에서 각 데이터의 시퀀스 데이터를 생성하고 저장한다. (line 5 - line9)에서 LSTM cell, layer의 개수를 활용해 LSTM 모델을 구축하고 충분히 높은 정확도를 얻을 때까지 모델을 학습하고 업데이트한다. 특히 (line 7)에서 Loss Function으로 cross-entropy를 선택한 이유는 multi-class 분류가 가능하기 때문이며 Optimizer로 Adam Optimizer를 선택한 이유는 기울기의 지수 평균을 저장하고, 기울기의 제곱의 지수 평균을 저장함으로써 기울기, 학습률을 빠른 시간 내에 최적화하는 것으로 알려졌다. 따라서 본 논문은 Adam Optimizer를 사용한다. (line 10 - line 11)에서 최적의 모델을 테스트하는 과정을 거친다.

4. 성능 평가

이동성 패턴을 미리 알고 이를 콘텐츠 캐싱에 적용한다면 cache storage의 사용률이 증가하기 때문에 cache hit ratio(HR)이 증가할 것이며 궁극적으로 energy efficiency가 높아진다. 다만 central server에서 머신러닝을 기반으로 이동성을 예측하기 때문에 computation power를 더 많이 사용하겠지만 이로써 content를 여러 SBSs에 걸쳐 빠르게 제공할 수 있으므로 SBSs의 service latency가 현저히 줄어들 것이며 결과적으로 network throughput도 증가할 것이다.

	Content Caching	Mobility-Aware Content Caching
Cache Hit Ratio	Normal	High
Energy Efficiency	Normal	High
Computation power of Central Server	Low	High
Service Latency	Normal	Low
Network Throughput	Normal	High

[표 1] Advantages of Mobility-Aware Content Caching

5. 결론 및 향후 연구

본 논문은 콘텐츠 캐싱을 위해 사용자의 이동 패턴을 예측하는 하나의 방법을 제시했다. 딥러닝 기반의 LSTM을 활용하여 사용자 이동성 데이터를 분석 및 예측하게 된다면 콘텐츠 캐싱뿐만 아니라 다양한 네트워크 자원을 효율적으로 배치하고 사용할 수 있게 될 것이다.

본 논문의 향후 연구로 이동 패턴 예측 accuracy를 높일 수 있는 mobile user의 이동성 예측에 중요한 영향을 끼치는 cell transition, cell sojourn time등을 feature로 실제 데이터를 기반으로 예측할 것이다. 또한 예측된 이동 패턴을 기반으로 사용자 군집(user association)을 형성하는 방법, 이들을 활용하여 콘텐츠를 최적으로 배

치하는 방법론 등을 연구함으로써 궁극적으로 Cache Hit Ratio(HR), Network Throughput 등을 극대화하는 것을 목표로 연구를 진행할 계획이다.

6. 참고문헌

- [1] R. Wang et al, "Mobility-Aware Caching for Content-Centric Wireless Networks: Modeling and Methodology," IEEE Communication Magazine, August 2016
- [2] X.peng et al., "Cache Size Allocation in Backhaul Limited Wireless Networks,"Proc. IEEE ICC, May 2016
- [3] Y. Ye et al, "Performance Analysis of Mobility Prediction Based Proactive Wireless Caching,"IEEE Wireless Communications and Networking Conference, 2018
- [4] T. Gao et al, "Reinforcement Learning based Resource Allocation in Cache-Enabled Small Cell Networks with Mobile Users,"IEEE/CIC International Conference on Communications in China(ICC), 2017
- [5] Abrahams et al, "Tensorflow for Machine Learning", Bleeding Edge Press, 2016