

Q-Cache: 대규모 무선 센싱을 위한 저지연 양자화 기반 캐시 기법

이동호, 김민서, 최재혁
가천대학교 AI·소프트웨어학부

{leedongho, mskim8055, jchoi}@gachon.ac.kr

Q-Cache: A Low Latency Quantization Based Caching for Large Scale Wireless Sensing

Dongho Lee, Minseo Kim, Jaehyuk Choi
School of Computing, Gachon University

요약

본 논문에서는 대규모 무선 센싱 환경에서 사람 행동 인식을 효율화 하기 위한 반복 패턴 중심의 캐시 구조 기반 효율화 기법을 제안한다. 반복적으로 발생하는 행동에 대한 전송 효율을 높이기 위해, 벡터 양자화 표현을 활용한 캐시 구조를 설계하였다. 제안한 구조는 반복 패턴을 식별하고, 이를 대표하는 표현 벡터를 로컬에 저장·활용함으로써 데이터 서버 전송 없이 추론을 가능하게 한다. 제안 시스템 구조는 전송량을 효과적으로 절감하면서도 높은 추론 정확도를 유지하여, 실시간성이 요구되는 에지 센싱 환경에 적합함을 확인하였다.

I. 서론

최근 스마트 홈, 병원, 산업 현장 등에서 수십에서 수천 개의 디바이스가 동시에 운용되는 대규모 무선 센싱 환경의 수요가 증가하고 있다 [1].

무선 신호 기반 행동 인식(Human Activity Recognition, HAR)은 Wi-Fi 채널 정보를 나타내는 채널 상태 정보(Channel State Information 이하 CSI)를 분석함으로써 이루어진다. CSI 는 송수신 안테나 간 주파수별 채널 응답을 복소수로 표현한 정보로, 환경 변화나 사람의 움직임에 민감하게 반응한다. 그러나 CSI 데이터는 고해상도 시계열로 구성되어, 전송량과 연산 비용이 매우 크다 [2]. 특히 대규모 센싱 환경에서는 이러한 데이터 송수신이 무선 채널의 과부하와 네트워크 병목 현상을 초래한다. 이를 완화하기 위해 해상도를 낮추거나 데이터 전송량을 줄일 경우, 행동 인식 정확도가 저하되는 트레이드 오프 문제가 나타난다.

본 연구에서는 이러한 한계를 극복하기 위해, 반복적으로 발생하는 행동 패턴을 효율적으로 처리하기 위한 캐시 기반 추론 구조를 제안한다. 양자화된 표현을 기반으로 클래스별 대표 패턴을 캐시에 저장함으로써, 반복 행동 에 대해 서버 전송 없이 로컬에서 근사 추론이 가능하도록 설계하였다. 이는 벡터 양자화(Vector Quantization 이하 VQ)의 특성을 활용한 방식으로, 대규모 센싱 환경에서 전송량과 연산량을 효과적으로 절감하면서도 인식 정확도를 유지할 수 있는 경량 추론 방식으로 기능한다.

II. 본론

2.1 기존 접근 방식 및 한계

기존의 무선 신호 기반 행동 인식 연구에서는 대규모 센싱 환경에서의 전송 부담을 줄이기 위해, VQ 기반의 인덱스 추론 구조를 활용하였다. 그림 1 과 같이, 고차원 CSI 특징은 인코더를 통해 추출된 후, 학습된 코드북을 기반으로 가장 가까운 코드북 벡터 위치에 대응되는 정수 인덱스로 양자화된다. 이때 코드북 크기 K 와 표현

차원 D 은 양자화된 인덱스의 표현력과 행동 인식 성능에 직접적인 영향을 미친다.

이러한 방식은 원본 CSI 특징 대신 인덱스만을 전송함으로써 전송량을 효과적으로 절감할 수 있다. 해당 인덱스는 서버로 전송되어 서버 측에서 코드북을 통해 원래 특징 벡터로 복원한 뒤 분류를 수행한다 [3]. 그러나 이 구조는 서버 측에서 수신된 인덱스를 다시 복원하여 분류를 수행해야 하므로, 반복적으로 발생하는 행동에 대해서도 서버에서 디코딩 및 분류 연산이 지속적으로 요구되어 서버 부하가 증가하는 한계가 존재한다.

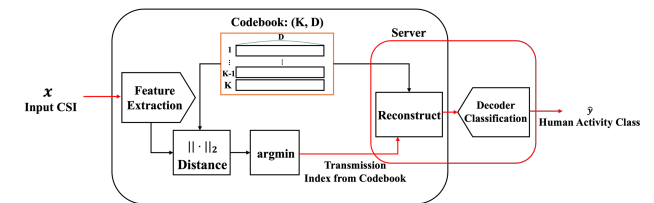


그림 1. 벡터 양자화 기반 추론 구조

2.2 제안 시스템 구조

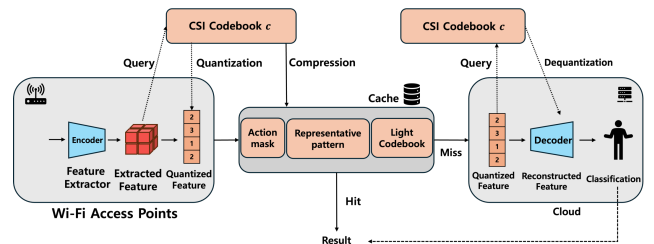


그림 2. 제안하는 캐시 기반 추론 구조

본 논문에서는 반복적으로 발생하는 행동에 대한 중복 연산과 전송을 줄이기 위해, 그림 2 와 같이 캐시 구조를 활용한 추론 구조를 제안한다. 본 시스템은 기존 VQ 기반 CSI 양자화 구조를 전제로 하여, VQ 기반 양자화 인덱스 생성, 학습 단계에서의 캐시 구성, 그리고 온라인 환경에서의 캐시 기반 추론 단계로 구성된다.

오프라인 단계에서는 캐시 기반 추론을 가능하게 하도록 학습 데이터로부터 클래스별 행동 패턴을 사전에 분석하여 캐시를 구성한다. 구체적으로, 클래스별 행동이 주로 발생하는 위치를 반영한 행동 마스크와 대표 인덱스 배열을 생성하며, 캐시 조회 시 연산량을 줄이기 위한 저차원 경량 코드북을 학습한다.

실제 추론 시에는 입력 CSI로부터 VQ 구조를 통해 생성된 정수 인덱스에 행동 마스크를 적용하여 중요 위치의 인덱스만을 선택적으로 추출한 뒤, 저차원 코드북 공간에서 복원된 벡터 표현을 기반으로 클래스별 대표 인덱스 배열과 코사인 거리 비교를 수행한다.

계산된 클래스별 거릿값 중 최솟값이 사전에 정의된 임계 값 이하일 경우, 해당 입력은 캐시 히트로 판단하여 로컬에서 즉시 행동을 예측하며, 임계 값을 초과하면 캐시 미스로 간주하여 서버로 인덱스를 전송해 복원과정을 통해 분류를 진행한다.

2.3 실험 환경 및 평가 지표

제안한 캐시 기반 추론 구조의 효율성과 설계 요소의 영향을 분석하기 위해 Widar3.0 데이터 세트를 활용하여 실험을 진행하였으며, Push & Pull, Sweep, Clap, Slide, Draw-N, Draw-O 총 6 가지 활동을 대상으로 했다 [4].

평가 항목은 크게 두 가지로 구성된다. 첫째, 코드북 벡터 차원에 따른 캐시 hit 비율과 추론 정확도 간의 상충관계를 분석하고, 둘째, 캐시 미스의 경우 발생하는 서버 전송량을 정량화 하여 코드북 설계에 따른 통신 효율성과 인식 정확도를 확인하였다.

2.4 실험 결과

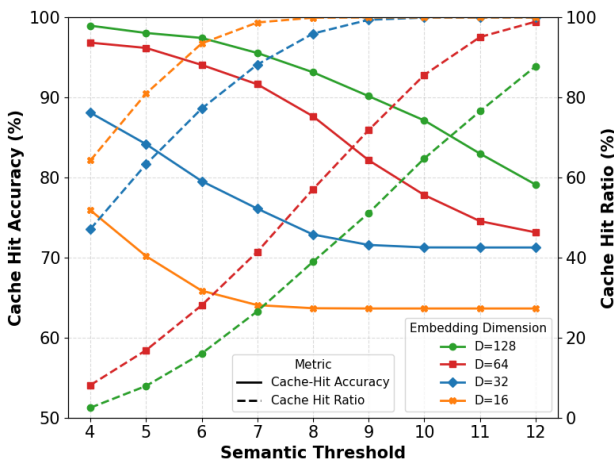


그림 3. 캐시 정확도와 캐시 hit 간의 트레이드오프

첫 번째 실험에서는 코드북 크기는 128 로 고정된 상태에서 벡터 차원을 변화시키며, 캐시 히트 상황에서의 추론 정확도와 캐시 적중률간의 관계를 분석하였다.

실험 결과, 벡터 차원이 증가할수록 캐시 기반 추론 정확도는 향상되었다. 이는 벡터 차원이 커질수록 코드북이 CSI 특징에 대한 더 풍부한 의미 정보를 보존할 수 있어, 캐시에 저장된 대표 인덱스 배열과의 유사도 매칭 정확도가 증가하기 때문이다.

또한, 제안한 캐시 구조는 반복적이고 특징적인 행동 패턴에 대해서만 높은 신뢰도로 캐시 히트를 발생시키며, 불확실한 샘플은 캐시 미스로 판단되어 서버로 전달된다. 이러한 선택적 처리 메커니즘은 캐시 기반 추론의 정확도를 유지하는데 기여한다.

두 번째 실험에서는 표 1 에 제시된 바와 같이, 코드북 크기와 벡터 차원에 따른 전송률과 정확도를 비교하였다. 기준 전송량은 Widar 3.0 환경의 원시 전송률로

설정하였으며, 이는 3 개의 안테나 쌍, 40 MHz 대역폭, 500Hz 샘플링 조건에서 약 4010 KB/s 에 해당한다.

Method	Transmission Rate (KB/s)	Cache Hit Rate (%)	Accuracy (%)
Widar 3.0 [5]	4010	–	98.9
EfficientFi [4]	48.00	–	98.8
Q-Cache (K=128, D=128)	17.20	68.5	93.2
Q-Cache (K=128, D=64)	20.95	60.2	93.0
Q-Cache (K=128, D=32)	17.90	66.4	90.8
Q-Cache (K=64, D=32)	23.42	53.9	92.1

표 1. 코드북 크기 및 벡터 차원에 따른 성능 변화 비교

실험 결과, 제안한 캐시 기반 시스템은 기존 서버 송신 방식 대비 전송량을 크게 절감하였다. 특히, 동일한 벡터 양자화 기반 전송 시스템과 비교했을 때도, 반복 행동 패턴에 대한 캐시 활용을 통해 전송량을 절반 이상 추가로 절감하는 효과를 보였다.

코드북 크기와 벡터 차원이 증가할수록 캐시 적중률은 향상되는 경향을 보였으나, 양자화 및 거리 계산 과정에서의 연산 복잡도와 실시간 추론 지연을 함께 고려할 때, 코드북 크기 128, 벡터 차원 64 의 조합이 전송 효율성과 인식 정확도 측면에서 가장 균형 잡힌 성능을 나타냄을 확인하였다.

III. 결론

본 연구는 대규모 무선 센싱 환경에서 발생하는 데이터 전송 병목 문제를 완화하기 위해, 벡터 양자화 기반 캐시 구조를 활용한 행동 인식 기법을 제안하였다. 제안 시스템은 양자화된 인덱스 표현과 캐시 기반 근사 추론을 결합함으로써, 반복적으로 발생하는 행동에 대해 서버와의 통신 없이도 로컬에서 신뢰도 높은 추론이 가능함을 실험적으로 확인하였다.

실험 결과, 제안한 캐시 기반 접근법은 기존 양자화 기반 전송 구조 대비 전송량을 절반 이상 절감하면서도 인식 정확도를 유지하였다. 이는 반복성이 높은 무선 센싱 환경에서 통신 비용과 서버 연산 부담을 동시에 완화할 수 있음을 보여주며, 대규모 무선 센싱 환경에서 실용적인 대안으로 활용될 수 있다.

ACKNOWLEDGEMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No.RS-2023-00252039). 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 2025 년도 SW 중심대학사업의 결과로 수행되었음 (2021-0-01389).

참 고 문 헌

- [1] Wei, Zhongchengm et al. "A survey on Wi-Fi based human identification: Scenarios, Challenges, and current solutions." *ACM Transactions on Sensor Networks* 21.1 (2025):1-32.
- [2] Barahimi, Borna, et al. "Rscnet: Dynamic csi compression for cloud-based wifi sensing." *ICC 2024-IEEE International Conference on Communications*. IEEE, 2024.
- [3] Yang, Jianfei, et al. "EfficientFi: Toward large-scale lightweight WiFi Sensing via CSI compression." *IEEE Internet of Things Journal* 9.15 (2022): 13086-13095.
- [4] Zhang, Yi, et al. "Widar3.0: Zero-effort cross-domain gesture recognition with Wi-Fi." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.11 (2021): 8671-8688.