

실내 위치 추정을 위한 우도 기반 동적 데이터 가지치기

이서인, 문하은, 한승우, 김민석*
충남대학교, *충남대학교

sslynn22@o.cnu.ac.kr, cherishuu02@o.cnu.ac.kr, hsw2266@g.cnu.ac.kr, * kms4105@cnu.ac.kr

Likelihood based Dynamic Data Pruning for Indoor Localization

Lee Seo In, Moon Ha Eun, Han Seung Woo, and Kim Min Seuk*
Chungnam National Univ., * Chungnam National Univ.

요약

본 논문은 Wi-Fi 채널 상태 정보(Channel State Information, CSI) 기반 실내 위치 추정 문제에서 실내 환경의 다중경로(Multipath) 및 NLoS(Non-Line of Sight) 등으로 유발되는 잡음·이상치로 인한 성능 저하 문제를 해결하기 위해, 학습 과정에서 샘플의 신뢰도를 실시간으로 평가해 유효하지 않은 데이터를 일시적으로 배제하는 우도(Likelihood) 기반 동적 데이터 가지치기를 제안한다. 실험 결과, 제안 기법은 기존 합성 신경망(Convolutional Neural Network, CNN) 모델(3.99 m) 대비 약 13 % 향상된 3.47 m 의 평균 위치 추정 오차를 달성하였다. 이는 학습 전 데이터의 정적 선별 방식이 지닌 한계를 극복하고, 모델의 강건성과 실내 위치 인식 정확도를 효과적으로 향상시킬 수 있음을 입증한다.

I. 서론

Wi-Fi 채널 상태 정보(Channel State Information, CSI)를 활용한 실내 위치 추정은 스마트 팩토리, 자율주행 등 다양한 분야에서 핵심 기술로 사용되고 있다. 그러나 다중경로(Multipath), NLoS(Non-Line of Sight) 등이 발생하는 복잡한 실내 환경에서 수집된 CSI 데이터는 잡음(Noise)과 이상치(Outlier)가 섞이게 된다 [1]. 이렇게 신뢰도가 낮은 데이터는 위치 추정 모델의 정확도와 일반화 성능을 저해하는 원인이 된다.

이러한 비선형적인 신호 패턴을 효과적으로 학습하기 위해 합성곱 신경망(Convolutional Neural Network, CNN)이 활용되고 있다 [2]. CNN 은 모든 데이터를 동일 가중으로 학습하기 때문에 잡음·이상치가 포함된 데이터의 영향을 그대로 받아, 위치 추정 성능의 한계를 보인다. 이를 완화하고자 전체 데이터를 무선별적으로 학습하는 기존 방식에서 벗어나, 신뢰도가 높은 핵심 데이터만을 선별하여 학습하는 우도(Likelihood) 기반 동적 데이터 가지치기(Dynamic Data Pruning) 모델을 제안한다. 모델의 학습 과정에서 각 데이터 샘플에 대한 예측 손실을 우도의 역지표로 정의하여 손실값이 큰, 즉 우도가 낮은 샘플을 실시간으로 제거한다. 이를 통해 모델의 학습 단계에서 신뢰도가 높은 데이터만을 선별하여 학습하므로, 강건성을 확보하고 최종 위치 추정 정확도의 향상을 달성할 수 있음을 보인다.

II. 본론

2.1 모델 아키텍처

본 논문에서 제안하는 우도 기반 동적 데이터 가지치기는 모델 학습 과정에서 현재 모델 상태를 반영하여 우도가 높은 샘플만 선택하고 나머지는 일시적으로 제외한다. 전체 프로세스는 (가) 우도 기반

샘플 신뢰도 산출과 (나) 동적 데이터 가지치기의 두 단계로 구성된다.

(가) 우도 기반 샘플 신뢰도 산출

데이터 가지치기의 판별 기준으로 각 샘플(X_i) 우도를 활용하며, 이를 정량화하기 위해 모델의 평균 제곱 오차(Mean Squared Error, MSE)를 우도의 역지표로 사용한다. 이는 학습 과정에서 실시간으로 데이터의 신뢰도를 평가하여, 모델 성능에 악영향을 미치는 잡음 및 이상치를 동적으로 학습에서 배제하기 위함이다.

$$L_i = L_{MSE}(f_{\theta}(Rep(X_i)), Y) \quad (i = 1, \dots, N) \quad (1)$$

현재 모델 파라미터 θ 를 기준으로 배치 내 개별 샘플에 대한 손실 값 L_i 를 수식 1 과 같이 산출한다. CNN 모델이 M 의 차원을 가지는 개별 샘플(X_i) 각각의 신뢰도를 평가할 수 있도록, 단일 데이터 샘플은 복제 연산($Rep(X_i)$)을 통해 N 번 복제되어 모델의 2 차원 입력 형식($M \times N$)에 맞는 텐서 형태로 변환된다. 이 복제된 텐서는 CNN 모델 f_{θ} 의 입력으로 사용되어 예측 좌표($f_{\theta}(Rep(X_i))$)를 출력하며, 실제 정답 좌표 Y 와의 MSE를 계산하여 해당 샘플의 손실값 L_i 를 산출한다.

$$S_L(X_i) = \exp(-L_i) \quad (2)$$

수식 2 와 같이, 산출된 손실값 L_i 를 $\exp$ 지수 함수에 매핑하여 최종 우도 값($S_L(X_i)$)을 계산한다. 이 변환을 통해 손실이 작을수록 $S_L(X_i)$ 은 1에 가까워지고, 손실이 클수록 0에 수렴하여 각 샘플의 신뢰도를 나타내는 최종 지표가 된다.

(나) 동적 데이터 가지치기

우도 값($S_L(X_i)$)을 기준으로, 우도가 낮은 샘플들을 실시간으로 학습에서 제거(pruning)한다.

$$S = \{X_i \mid rank_{desc}(S_L(X_i)) \leq K\} \quad (3)$$

선별된 샘플의 집합 S 는 수식 3 과 같이, 전체 샘플(X_i)중에서 해당 우도 값($S_L(X_i)$)의 순위가 K 이하인

샘플들로 정의한다. 여기서 $rank_{desc}$ 는 값을 내림차순(Descending)으로 정렬했을 때의 순위를 의미하며, 이는 우도가 가장 높은 K 개의 샘플을 선택하는 것을 뜻한다. 이때, 전체 샘플 수(N) 대비 선택되는 샘플 수(K)의 비율, 즉 가지치기 비율(K/N)은 모델이 얼마만큼의 데이터를 선별할지를 결정하는 핵심 파라미터가 된다. 본 실험에서는 전체 샘플 그룹(N)을 10 으로, 선택할 샘플 수(K)를 5 로 설정하여, 매 학습 단계마다 신뢰도가 낮은 하위 50 % 의 샘플을 가지치기하였다.

데이터 가지치기 후, 선택된 K 개의 핵심 샘플은 다시 원래 배치 크기인 N 개를 유지하도록 복제되어 학습에 사용된다. 이는 학습 과정에서 CNN 모델의 입력 텐서 크기를 일정하게 유지하여 학습 안정성을 높이기 위함이다. 이렇게 선별된 핵심 샘플 집합 S 만을 사용하여 모델의 파라미터 θ 를 업데이트한다. 역전파(Backpropagation) 과정에서 사용되는 그래디언트(Gradient)가 신뢰도 높은 데이터로부터만 계산되므로, 잡음 및 이상치가 파라미터 업데이트에 미치는 악영향을 효과적으로 차단할 수 있다. 또한 학습이 반복될수록 모델의 성능이 향상되고, 향상된 모델은 다시 다음 Epoch 에서 더 정확하고 신뢰도 높은 데이터를 선별하는 선순환 구조를 형성하게 된다.

2.2 실험 환경 및 성능 평가

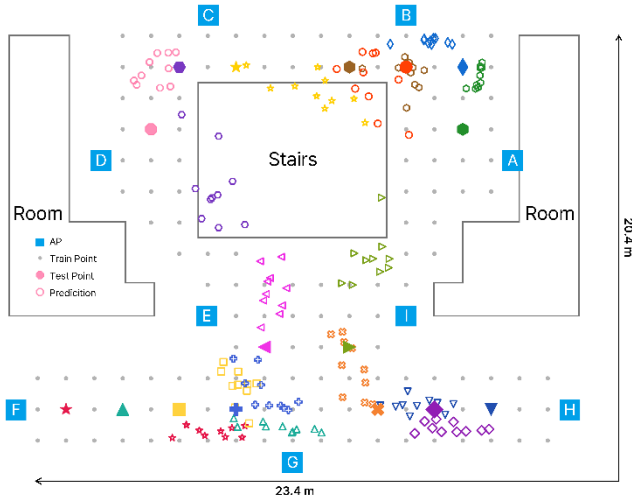


그림 1. 실험 환경 토폴로지 및 성능 평가 결과

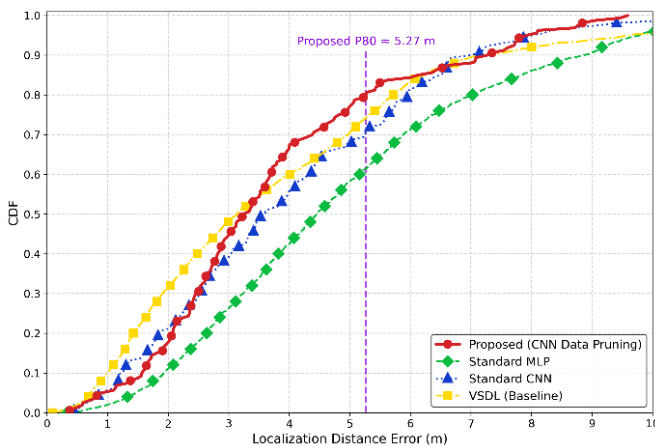


그림 2. 위치 추정 오차 CDF

본 논문에서 제안하는 모델의 성능은 그림 1 과 같이 유동 인구가 많은 충남대학교 도서관 1 층 실내 환경에서 검증하였다. 평가에는 총 141 개 지점에서 포인트당 100 회 측정된 1,890 차원 CSI 기반 특징 벡터를 사용하였다. 각 지점의 100 개 패킷은 10 개의 연속 구간으로 묶어

10 개의 이미지 청크로 변환하였고, 각 청크는 (10×1,890) 행렬로 CNN 에서 하나의 이미지처럼 처리되도록 구성하였다. 이로써 모델이 짧은 시간 구간의 변화(10 패킷)와 주파수·공간적 특성(1,890 차원 특징)을 동시에 학습할 수 있도록 하였다. 그림 1 에서 볼 수 있듯이, 테스트 포인트에 대한 실내 위치 추정 결과가 안정적으로 실제 위치에 자리 잡고 있다.

그림 2 는 타 알고리즘과 비교하여, 제안한 방법의 실내 위치 인식 정확도를 보여준다. 제안한 우도 기반 동적 데이터 가지치기는 동일 조건에서 평균 위치 추정 오차 3.47 m 로, 비교 모델인 MLP 4.65 m, CNN 기반 CiFi 3.99 m, 뷰 선별 알고리즘을 제안하는 VSDL [3] 4.01 m 대비 향상된 위치 추정 정확도를 기록하였다. 특히, 동일한 CNN 기반의 CiFi 모델(3.99 m)과 비교했을 때, 모든 데이터를 학습하는 대신 우도가 높은 데이터만을 선별하여 학습함으로써 오차율을 약 13 % 개선하는 효과를 확인했다. 그림 2 와 같이 P80(상위 80 % 보장 오차)가 약 5.27 m 로 나타나, 동일한 보장수준에서 타 방법들보다 더 낮은 임계 오차를 제공하였다.

III. 결론

본 연구는 Wi-Fi CSI 기반 실내 위치 추정에서 잡음·이상치로 인한 성능 저하 문제를 해결하고자, 학습 과정에서 데이터 신뢰도를 실시간으로 평가하고 유효하지 않은 샘플을 동적으로 배제하는 우도 기반 데이터 가지치기 기법을 제안했다. 실험 결과, 제안 기법은 기존 CNN 모델(3.99 m) 대비 약 13 % 향상된 3.47 m 의 평균 위치 추정 오차를 달성하였다. 이는 데이터의 정적 선별 방식이 지닌 한계를 극복하고, 모델의 강건성과 위치 추정 정확도를 효과적으로 향상할 수 있음을 입증한다.

향후 연구로는 본 연구의 샘플 평가 개념을 추론 단계까지 확장하여, 신뢰도 기반의 가중 평균이나 앙상블 기법을 통해 예측 안정성을 높이는 방안을 모색할 것이다. 또한, 고정된 가지치기 비율을 학습 상황에 따라 동적으로 조절하여 학습 효율을 최적화하는 연구를 진행할 계획이다.

ACKNOWLEDGMENT

본 연구는 충남대학교 학술연구비에 의해 지원되었음.

참 고 문 헌

- [1] Tzur, Asaf, Ofer Amrani, and Avishai Wool. "Direction finding of rogue Wi-Fi access points using an off-the-shelf MIMO-OFDM receiver." *Physical Communication* 17 (2015): 149–164.
- [2] Wang, X., Wang, X., and Mao, S. "CiFi: Deep convolutional neural networks for indoor localization with 5 GHz Wi-Fi," in *IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.
- [3] 한승우, 문하은, 이서인, 강민서, and 김민석, "Viewpoint-Selective Deep Learning 을 활용한 NLoS 환경에서의 Wi-Fi CSI 기반 실내 측위," in *한국통신학회 학술대회논문집*, 2025, pp. 1045–1046.