

PPO 를 활용한 실내 측위를 위한 향상된 AP 배치 기법

황준규, 동자빈, 릴라 라나, 리진롱, 박준구*
경북대학교 전자전기공학부

{junkyu891015, dong116745, ranaankit1171}@gmail.com, {2023000464, jgpark}@knu.ac.kr*

An Enhanced Access Points Placement Method For Indoor Positioning Using PPO

Hwang Jun Gyu, Jiabin Dong, Lila Rana, JinLong Li, Park Joon Goo*
Kyungpook National University, School of Electronic & Electrical Engineering

요약

위치기반서비스(LBS : Location Based Service)가 일상적으로 제공되고 있지만 복잡한 환경에서 높은 측위 성능이 요구되고 있다. 그래서 다양한 측위 기법 연구가 이루어지고 있다. 이에 본 논문은 PPO(Proximal Policy Optimization)을 활용한 실내 측위를 위한 AP(Access Point) 배치 기법을 제안한다. 강화학습기법인 PPO 를 적용하기 위해 RSSI(Received Signal Strength Indicator)와 DoP(Dilution of Precision)로 평가함수를 설계했다. 이를 통해 다양한 실내환경에서 측위 성능이 향상된 AP 들의 위치를 구한다.

I. 서론

무선통신기술의 발전으로 인해 위치기반서비스는 많은 분야에 적용되고 있다. 하지만 다양한 서비스를 위해서는 지금보다 높은 측위 성능을 가진 기술들이 필요하다.

이를 해결하기 위해 다양한 측위 기법들이 연구되고 있다.[1] 또, 인공지능기술 활용이 대두되며 이를 활용한 연구도 활발히 진행되고 있다.[2]

측위 성능 향상을 위해 인공지능기술은 크게 3 가지 파트에서 적용된다. 첫째로는 측정된 데이터(RSSI, RTT, AoA 등)의 정확성을 높이기 위해 사용된다. 둘째로는 측정된 데이터들을 입력하여 위치를 출력하는 네트워크를 만드는 데 사용되고 마지막으로 측위 환경을 분석 및 개선에 사용된다.

본 논문에서는 RSSI 와 DoP 로 설계된 측위 성능 평가함수로 PPO 기법을 사용하여 향상된 AP 들의 위치를 구하는 기법을 제안한다.

II. RSSI & DoP 가치함수 및 PPO

강화학습은 주어진 환경에서 설정된 행동들을 반복하며 보상을 많이 받는 방향으로 파라미터들을 수정하며 학습을 진행한다. 이런 반복학습에서 보상을 정하는 가치함수는 강화학습을 특정분야에 적용할 때 그 특성에 맞게 설계를 해야 의미 있는 결과를 도출한다.

이에 실내 측위 정확도를 나타낼 수 있는 RSSI 와 DoP 를 선택했다. RSSI 는 수신신호강도로 신호의 신뢰도와 거리를 나타내는 중요한 지표이다. RSSI 는 크기가 클수록 신뢰도가 높으며 거리가 가까움을 나타낸다. 또 RSSI 는 dBm 단위로 측정되며 $-15 \sim -100dBm$ 의 값을 가진다. DoP 는 위성항법에서 사용하는 지표로 위성들의 상대위치에 따른 측위 성능을 나타내는 지표이다. DoP 는 낮아질수록 오차범위가 줄어들어 측위 성능이 높아진다. 고려하는 좌표계에 따라 Horizontal DoP, Vertical DoP, Position(3D) DoP, Geometric DoP 로 나뉜다. 이런 특성들을 고려했을 때 가치함수 P 는 식(1)과 같이 제안한다.

$$P(s_{RSSI, DoP}) = \frac{1}{m} \sum_{i=1}^m (RSSI_i \times DoP_i)$$

$$RSSI_i = \sum_{j=1}^n |RSSI_j|$$

$$DoP = \text{trace}((A^T A)^{-1})$$
$$A = \begin{bmatrix} x_1 - x & y_1 - y & -1 \\ d_1 & d_1 & -1 \\ \vdots & \vdots & \vdots \\ x_j - x & y_j - y & -1 \\ d_j & d_j & -1 \end{bmatrix}$$

$$d_j = \sqrt{(x_j - x)^2 + (y_j - y)^2 + (z_j - z)^2}$$

x, y, z : 참조지점
 x_j, y_j, z_j : AP 의 위치
(1)

$RSSI_i$ 와 DoP_i 는 i 번째 참조지점(RP; Reference Point)에서의 RSSI 와 DoP 를 말한다. 참조지점은 환경 내에서 $1m \times 1m$ 공간으로 나눈 지점들을 나타낸다. 한 참조지점에서 n 개의 AP 들로부터 데이터를 수집할 때 RSSI는 AP 개수만큼 받은 후 절댓값의 평균으로 설계했고 DoP는 n 개의 AP 들의 배치 위치와 참조지점을 기준으로 계산하였다. 그림 1.에서는 위치오류와 가치함수의 상관도를 보여준다.

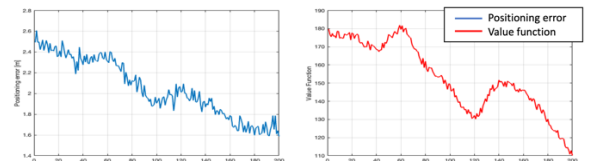


그림 1. 측위오차와 가치함수

PPO 알고리즘은 TRPO(Trust Region Policy Optimization) 알고리즘을 발전시킨 방법으로 정책 업데이트시 발산 할 수 있는 TRPO 의 단점을 보완하기 위해 정책 업데이트를 제한하는 클리핑 함수를 사용하여 새

정책이 이전 정책에서 너무 멀리 벗어나지 않도록 한다.[3] 이 특성을 통해 제한된 환경에서 AP 들이 나가지 않도록 하기 용이하다. 본 연구에서는 강화학습의 Action 으로 AP 의 위치 변화, Reward 는 측위 성능 향상, State 는 AP 의 위치와 환경요소들을 사용하였다.

PPO 정책 π_{θ} 를 초기화하는 과정, 에포크마다 환경을 리셋하고 각 타임스텝에 대해 정책 π_{θ} 를 사용하여 AP 의 위치변화를 취할지 선택하는 과정, 그리고 보상과 새로운 상태를 관찰하여 정책을 업데이트 하는 과정으로 나눌 수 있다. PPO 클리핑 매개변수 ϵ 은, 학습률을 의미하며 정책의 변화를 제한한다. $r(\theta)$ 는 새 정책과 이전 정책의 비율로 새 정책에서 특정 행동을 할 확률을 이전 정책에서 했을 확률로 나눈 값이다. clip 함수는 $r(\theta)A_t$ 을 $1-\epsilon$ 과 $1+\epsilon$ 사이로 제한하여, 정책의 급변을 방지하고 학습 안정도를 높인다. r_t 는 시간에 따라 에이전트가 받은 보상을, s_t 는 시간에 따른 환경 상태를, a_t 는 시간에 따라 에이전트가 취한 행동을 나타낸다. A_t 는 시간에 따른 이득을 나타낸다. 결론적으로 A_t 가 가장 높은 값을 갖도록 s_t 의 변화에 대한 a_t 가 효율적이 되도록 학습된다.

III. 시뮬레이션 및 결과

시뮬레이션 환경은 30m x 30m 공간으로 가정하였다. PPO 을 통한 AP 위치조정은 Python 을 통해 진행하였고 Fingerprint-Map 을 이용한 측위 시뮬레이션은 Matlab 을 사용하였다. 대상 공간을 그림 2.(a-1)과 같이 환경 1, 그림 6.(b-1)과 같이 환경 2 로 구분하였다.

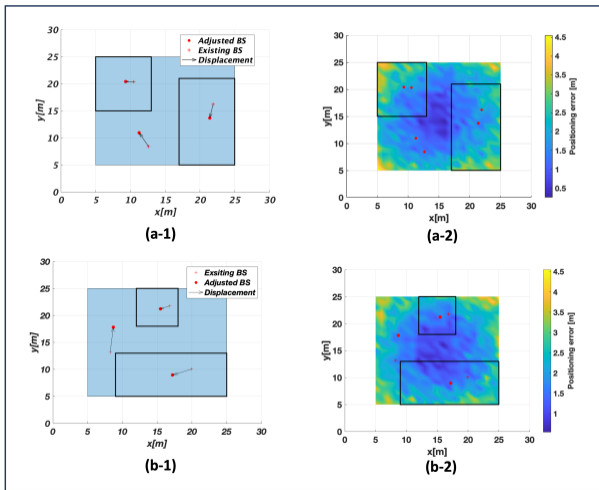


그림 2. 시뮬레이션 결과

그림 2.(a-1), (b-1)에서 측위성능을 높일 수 있는 기지국의 위치는 세부구역(방)의 중심으로 조정되는 것으로 확인된다. 그림 2.(a-2), (b-2)는 모든 참조지점의 측위성능을 나타낸다. 전체적으로 평균 측위 성능이 기존방법보다 0.2m 향상되었다.

표 1. 기존 방법과 제안 방법 성능 비교

		기존방법	제안방법
환경1	평균측위오차 [m]	1.62	1.49
	평균측위오차 [m]	1.79	1.52

IV. 결론

본 논문에서 실내측위정확도 향상을 위한 AP 배치 기법을 제안했다. PPO 기법을 기지국배치에 적용하기 위해 가치함수를 RSSI 와 DoP 를 사용하여 설계했다. 제안한 기지국배치 방법은 제시한 환경에서 모든 참조지점들의 평균측위오차를 기존방법보다 약 0.2m 줄였으며, 12.5% 측위정확도를 향상시켰다. 향후 개선된 가치함수 설계 및 환경요소들을 입력하면 최적기지국 위치가 출력으로 나오는 함수(Net)를 연구할 예정이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 지역지능화혁신인재양성사업임 (IITP-2024-RS-2022-00156389).

참고 문헌

- [1] Ji, Yufeng, et al. "Generating indoor Wi-Fi fingerprint map based on crowdsourcing." *Wireless Networks* 28.3 (2022): 1053-1065.
- [2] Jayakanth K, Comparative analysis of texture features and deep learning method for real-time indoor object recognition. *International Conference on Communication and Electronics Systems (ICCES)*, IEEE, pp 1676-1682, 2019.
- [3] SCHULMAN, John, et al. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.