

무선 센서 네트워크에서 강화학습 기반의 동적 듀티 사이클 모드 선택 방법

윤성호*, 최수연*, 김소명*, 윤완규*, 최승희*, 유상조*

A Reinforcement Learning Based Dynamic Duty-Cycle Mode Selection Method in Wireless Sensor Networks

Seong-Ho Youn*, Sue-Yeon Choi*, So-Myeong Kim*,
Wan-Kyu Yun*, Seung-Hee Choi*, Sang-Jo Yoo*

요약

무선 센서 네트워크 환경에서 센서 노드들은 지속적으로 데이터를 수집해 싱크로 전송하여 환경을 실시간으로 모니터링 한다. 이러한 환경의 센서는 자원에 한계가 있으므로 에너지를 효율적으로 사용하는 것이 중요하다. 또한, 무선 센서 네트워크에서 객체 추적을 할 때 객체 추적 정확도 역시 중요한 요구사항이다. 두 가지 사항을 모두 높은 수준으로 만족시키기 위해 본 논문에서는 강화학습 기반으로 객체의 미래 이동을 예측하여 센싱 모드를 동적으로 전환하는 방안을 제안한다. 강화학습을 위하여 센서 노드들에서 싱크로 전송된 센싱 데이터를 이용해 객체의 속력과 방향을 나타내는 현재 상태를 정의한다. 이를 Q-learning에 사용하여 센서 영역의 센서들을 각 상태에 맞는 최적의 모드로 전환시킨다. 모의실험을 통해 제안된 방법이 객체 추적 시 높은 정확도를 만족한 상태에서 에너지 효율성도 증대시키는 것을 확인하였다.

Key Words : Wireless Sensor Network, duty cycle, Q-learning, dynamic scheduling

ABSTRACT

Wireless Sensor Network monitors the environment in real-time by continuously collecting data from sensor nodes and sending it to the sync. Sensors have limited resources, so it is important to use energy efficiently. In addition, object tracking accuracy is also an important requirement when object tracking is performed in wireless sensor networks. To satisfy both at a high level, this paper proposes to dynamically switch sensing modes by predicting future movements of objects. Once the sensor nodes have synchronized sensing data to define the current state of object speed and direction. We use Q-learning to put it into a wake-up mode that temporarily leaves the optimal sensor region for each state on. Through simulations, we confirm that the proposed method increases energy efficiency with certain levels of accuracy satisfied.

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2021-0-02052)

♦ First Author : Inha University Department of Information and Communication Engineering, grympan@naver.com, 학생회원

° Corresponding Author : Inha University Department of Information and Communication Engineering, sjyoo@inha.ac.kr, 종신회원

* Inha University

논문번호 : 202107-163-B-RN, Received July 12, 2021; Revised August 17, 2021; Accepted September 25, 2021

1. 서 론

무선 센서 네트워크(WSN : Wireless Sensor Network)는 센서 노드들 간 애드혹 네트워크를 구성하여 기지국 없이 무선으로 네트워크를 구성하고 데이터를 수집 및 교환하는 기술이다. 이는 정부의 관심 속에서 꾸준히 연구되어 왔으며 자연재해 감시, 홈 네트워크, 재고 관리 등에 활용되고 있다. 무선 센서 네트워크에서 센서 노드들은 배터리로 동작하기 때문에 배터리의 수명이 다하면 데이터를 수집할 수 없다.^[1,2] 그러므로 무선 센서 네트워크 설계 시 가장 먼저 고려되어야 할 사항은 각 노드들의 에너지 소모를 줄여 센서 네트워크의 전체적인 수명을 늘리는 것이다.^[3] 센서 네트워크를 활용한 다양한 응용 연구 중에서 이동 객체 추적 기법은 활발하게 연구되고 있는 응용 연구 중 하나이다.^[4] 무선 센서 네트워크에서 에너지 소모를 줄이고 높은 정확도로 객체를 추적하기 위해 많은 연구가 진행되고 있다. Juan Luo는 실내의 무선 센서 네트워크에서 안정적으로 객체를 추적하기 위한 cooperative localization and tracking algorithm (CLTA)를 제안했다.^[5] 두 개 이상의 네트워크가 있는 실내 환경에서 서로 다른 네트워크의 특성을 고려하고 협력하여 객체의 위치를 높은 정확도로 추적하였다. S.W. Jang은 연속적인 객체를 에너지 효율적으로 추적하고, 감지 경계 노드의 수를 줄일 뿐 아니라 연속적인 객체의 형태를 가장 근사하게 표현할 수 있도록 지역적인 드로네 삼각 기법을 이용하였다.^[6] 실험을 통해 통신 메시지 수와 에너지 효율성이 향상됨을 보였다. Bo Jiang는 단일 객체 추적에서 에너지 효율성 개선을 위한 PPSS(Probability-based Prediction and Sleep Scheduling protocol)를 제시했다.^[7] 이는 운동학과 확률을 기반으로 객체의 다음 위치를 예측할 뿐만 아니라 모든 방향을 따라 이동하는 확률을 설명하고 이를 기반으로 활성화할 노드들을 선택하며 활성 시간을 제어하여 에너지 효율성을 향상시킨다. 본 논문에서는 센서 노드의 에너지 소모를 줄이기 위해 센서가 듀티 사이클(duty cycle)에 따라 동작하는 네트워크를 고려한다. 듀티 사이클 방식은 네트워크의 센서가 슬립 상태와 웨이크업 상태를 주기적으로 반복하도록 동작시키는 것을 의미한다.^[8] 이 때, 슬립 상태의 시간이 길어지면 에너지 소모가 줄어들어 장점을 가지지만 반대로 데이터 수집량이 줄어들게 되어 정확한 실시간 데이터를 싱크에게 제공하지 못하는 문제가 발생하게 된다.^[9] 이에 센서 노드들이 최적의 듀티 사이클 주기를 설정할 수 있도록 듀티 사이

클 스케줄링 기법에 대한 다양한 연구가 진행되고 있다. S.W. Lee는 모든 센서 노드의 잔여 에너지양과 수집 가능한 에너지양을 함께 고려하여 사용 가능한 총 에너지양에 따른 듀티 사이클 주기의 변화량을 서로 다른 방식으로 조절하는 세 가지 듀티 사이클 스케줄링 기법을 제안했다.^[10] H.K. Lee는 듀티 사이클 기반의 무선 센서 네트워크에서 에너지 효율성을 높이고 시간 지연을 줄일 수 있는 경로설정을 위해 듀티 사이클을 고려한 라우팅 매트릭 EDW을 고안하고 이를 활용한 크로스 레이어 기반의 다중경로 라우팅 프로토콜을 제안했다.^[11] 실험을 통해 낮은 전송 지연 시간과 적은 에너지 소비를 확인하였다. M.H. Kim은 빌딩 자동화 및 홈 센서 네트워크에서 에너지 효율성을 고려하여 실제 이웃 노드 집합에 기반한 실용적인 듀티 사이클링 메커니즘을 제안했다.^[12] 주기적으로 센싱 값을 바탕으로 센싱, 전송, 수신을 수행하는 액티브 모드 센서 셋을 결정하고, 라우팅 정보와 함께 결정된 셋 정보를 알린다. 해당 액티브 모드 센서 셋에서 배제된 노드들은 다음 결정 주기까지 슬립 모드를 취하는 방식으로 에너지를 절약한다.

현재 듀티 사이클과 관련된 연구들은 에너지 효율성에만 초점을 맞춰 객체 추적의 정확도는 고려하지 않는다는 한계점이 있다. 강화학습 기법 중 하나인 Q-learning을 사용하여 객체가 이동할 범위를 사전에 예측하고 센서의 듀티 사이클 모드를 반복적으로 전환해 에너지 소모는 낮추고 객체를 안정적으로 추적할 수 있는 방법을 제안한다. 강화학습은 임의의 환경에서 학습 주체인 에이전트의 행동 결과에 대한 보상을 최대화하는 방향으로 최적의 행동을 학습하는 방식이다.^[13] Q-learning은 강화학습 방법들 중 대표적으로 쓰이는 알고리즘으로 주어진 환경과 상태에서부터 비지도 학습을 통해 최적값을 찾아가는 방식이다.^[14] 에이전트(agent)가 환경에 대한 사전 정보를 알지 못해도 적용할 수 있다는 장점을 가지고 있다.

논문의 구성은 다음과 같다. 2장에서는 본 논문이 제안하는 무선 센서 네트워크에서 강화학습 기반의 동적 듀티 사이클 모드 선택 방법에 대해서 설명한다. 3장에서는 가상의 시뮬레이션 환경을 구성하여 본 논문이 제안하는 방법에 대한 성능에 대해서 평가하며 마지막 4장에서 결론으로 논문을 마무리한다.

II. 강화학습을 이용한 센싱 모드 동작 제어 방법

본 논문이 제안하는 시스템 구조는 그림 1과 같다. 센서의 동작 방식으로 두 가지 듀티 사이클 모드를

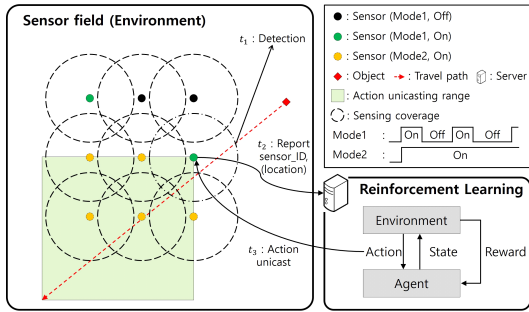


그림 1. 시스템 모델
Fig. 1. System model

사용한다^[15]. 모드 1은 슬립 상태와 웨이크업 상태를 반복하는 모드로 설정하고 모드 2는 웨이크업 상태를 지속하는 모드로 설정한다. 네트워크에서 객체의 이동을 감지한 센서들은 싱크로 센서의 위치 및 감지 시간을 전송한다. 평상시에는 센서의 에너지 소모를 줄이기 위해 모드 1을 사용하지만, 센서에서 객체가 감지되면 현재 상태를 정의하고 강화학습을 통해 각 상태에서 최적으로 선택된 범위의 센서들을 모드 2로 전환한다. 현재 상태는 싱크에서 센싱 데이터를 기반으로 객체의 움직임 경향성을 토대로 도출한다. 도출된 현재 상태를 강화학습에 이용하여 미래의 객체 이동 범위를 예측한다. 이동 범위에 대한 정보를 다시 센싱 데이터를 전송한 노드에게 전달하면 해당 노드는 범위에 속하는 센서들에게 유니 캐스트 방식을 사용하여 모드 2로 전환하도록 알린다. 실제 객체가 이동을 진행한 뒤 예측 결과를 평가하여 학습 모델의 정확도를 높여간다.

2.1 Q-learning

본 논문에서는 객체의 움직임 예측해 최적의 센싱 모드에 대해 실시간으로 수학적으로 도출하는 것이 쉽지 않다는 점을 고려해 센서의 센싱 모드 제어 방법으로 Q-learning을 사용한다. 강화학습은 행동을 평가하는 가치 함수(Q-Value)에 의해 양 또는 음의 보상을 부여함으로써 점진적으로 각 상태에 따른 최적의 행동을 찾아낸다.^[16] 그림 2와 같이 에이전트는 현재 상태 s_t 를 인지하고 각 상태에 따른 최적의 행동 a_t 을 한다. 환경은 행동의 결과로써 보상 r_t 을 반환한다.

Q-learning에서 학습 대상이 되는 것은 가치 함수이다. 이 값은 현재 상태에서 행동을 선택하기 위한 지표가 되는 수치이다. 이때 보상을 이용하여 Q값을 수정해 적절한 Q값을 얻는 것이 학습의 목표가 된다.

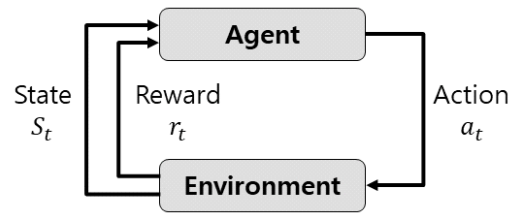


그림 2. 강화학습 모델
Fig. 2. Reinforcement learning model

가치 함수는 수식 (1)에 의해 갱신된다.

$$Q(s_t, a_t) = (1 - \alpha)Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a_t'} Q(s_t, a_t')] \quad (1)$$

수식에서 s_t, a_t 는 각각 현재 시점에서 상태와 행동이고 s_t, a_t' 는 미래 시점에서의 상태와 행동으로 다음 상태와 그에 상태에 대한 행동으로 정의한다. r_t 은 현재 상태를 기준으로 하여 다음 상태에 받게 될 것이라 기대되는 보상이다.^[17] α 는 (0, 1)사이의 값을 가지는 학습률로 이전의 값을 얼마만큼 반영할지를 나타내며 α 가 커질수록 새로운 정보를 많이 반영하게 된다. γ 는 미래의 보상 가중치를 의미한다.^[18] 결론적으로 현재 상태의 누적 보상은 현재 상태의 행동 값과 다음 상태의 보상과 다음 상태의 행동에서 가장 큰 값으로 구해진다.

학습의 정확도를 높이기 위해 최적의 행동을 선택할 때 decayed epsilon greedy 방식을 사용한다.^[19] 해당 방식은 학습 초기에는 대부분 무작위로 행동을 선택하도록 하여 여러 행동을 탐색하도록 하며, 후기에는 가치 함수가 최대인 행동을 선택하도록 한다. 두 가지 경우를 판단하는 경계 값은 $\epsilon_{threshold}$ 로 수식은 (2)와 같다. 무작위로 선택된 값이 $\epsilon_{threshold}$ 보다 크면 최대 가치 함수를 가진 행동을 선택하고 작으면 무작위로 행동을 선택한다. 과정은 그림 3에 나타낸 알고

Pseudo-Code for decayed epsilon greedy

```

1:  $\epsilon_{threshold} \leftarrow \epsilon_{end} + (\epsilon_{start} - \epsilon_{end}) \cdot e^{-\frac{t}{\tau}}$ 
//  $\epsilon_{threshold}$  denotes Boundary value that determines two methods with optimal action
2: for  $i = 1$  to  $M$  //  $M$  denotes the number of tasks to be scheduled
3:    $random = rand()$  // random number extraction
4:   if  $random > \epsilon_{threshold}$  then
5:     select action corresponding to max Q
6:   else
7:     select action at random
8:   end if
9: end for
    
```

그림 3. decayed epsilon greedy 과정
Fig. 3. decayed epsilon greedy pseudo code

리즘과 같다.

$$\epsilon_{threshold} = \epsilon_{end} + (\epsilon_{start} - \epsilon_{end}) \cdot e^{-\frac{i}{\epsilon_{decay}}} \quad (2)$$

수식에서 i 는 진행 횟수로 i 가 커질수록 $\epsilon_{threshold}$ 는 줄어들게 되며 최대인 가치 함수를 선택할 확률이 높아지게 된다. ϵ_{end} 는 마지막 epsilon 값이고 ϵ_{start} 는 초기 epsilon 값이다. ϵ_{decay} 는 감소하는 정도를 나타내며 값이 클수록 $\epsilon_{threshold}$ 가 크게 감소한다.

2.2 동적 센싱 모드 제어 방법

센싱 모드에는 웨이크업 상태와 슬립 상태를 반복하는 모드1과 항상 웨이크업 상태를 유지하는 모드2로 구분한다. 웨이크업 상태에서는 객체 감지가 가능하며 구동을 위해 많은 에너지를 소모하게 된다. 반면 슬립모드에서는 센서가 비활성화된 상태이기 때문에 에너지를 소모하지 않으며 객체 역시 감지할 수 없다. 따라서 모드1은 센서의 슬립 상태 구간에서 객체를 감지할 수 없어 감지 정확도는 하락할 수 있으나 에너지를 절감할 수 있다. 반면, 모드2는 객체를 항상 감지할 수 있어 정확도 측면에서는 가장 이상적인 모드인 동시에 에너지 소모가 가장 많은 모드이다. 제안하는 동적 센싱 모드 제어 방법은 센서 네트워크상의 모든 센서 노드는 모드1을 기본으로 동작하도록 하여 소모 에너지를 크게 절감하고, Q-learning을 이용해 객체의 예상 이동 경로에 위치한 센서는 모드2로 미리 변경하여 감지 정확도를 일정 수준으로 확보한다. 이를 통해 기존 듀티 사이클 프로토콜의 감지 정확도와 에너지 소모 간의 트레이드오프 관계의 격차를 최소화할 수 있다. 센싱 모드의 자연스러운 동작을 위해 두 가지 조건을 추가한다. 첫 번째는 모드1에서 객체를 감지하는 경우 해당 센서를 모드2로 즉시 전환하도록 설계하는 것이다. 두 번째는 Q-learning을 통해 모드2로 전환한 경우, 이로부터 T_{hold} 시간이 경과될 때까지 객체를 감지하지 못하면 모드1로 되돌아가도록 설정한다. 반면 그 전에 감지한 경우 객체가 감지 반경을 벗어나 감지되지 않을 때까지 모드2를 유지한다.

최적의 센싱 모드를 예측하기 위한 Q-Learning 모델의 환경 설정을 위해 우선 현재 상태를 정의한다. 현재 상태는 객체의 움직임을 나타내는 벡터로 속력과 방향을 나타내도록 설정한다. 센서들은 객체가 이동하는 것을 감지할 때마다 싱크에게 해당 센서의 위치와 감지시간을 전송한다. 싱크에서는 일정 시간동안

감지 데이터를 모으고 현 시점에 감지된 센서의 위치를 interest region의 중심으로 설정한다. 싱크로 전달된 데이터 중 일정 시간 이내에 해당 영역 안에서 감지된 데이터를 이용해 객체 움직임에 대한 벡터를 계산한다. 그림 4는 이 과정을 그림으로 나타낸 것이다. 데이터들 중 가장 과거에 감지된 센서의 좌표를 x_1, y_1 라 하고 가장 최근에 감지된 센서의 좌표를 x_2, y_2 라 한다. 두 좌표를 이용하여 현재 상태 벡터를 추정한다. 벡터의 좌표와 속도는 수식 (3),(4)에 따라 도출된다. 여기서 t_m 는 객체가 x_1, y_1 에서 x_2, y_2 로 가는데 걸린 시간이다.

$$\vec{v} = (x_2 - x_1, y_2 - y_1) \quad (3)$$

$$v = \frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{t_m} \quad (4)$$

Q-Learning의 상태를 너무 다양하게 정의하게 되면 학습에 어려움이 있기 때문에 방향과 근사화를 통해 그림 5와 같이 속도에 따라 17개의 상태로 구분한다. 수식 (5)를 통해 x축과 객체의 이동 방향 간의 각도를 계산하며 이를 범위에 따라 각각 상, 하, 좌, 우, 대각선 상태로 근사화 한다.

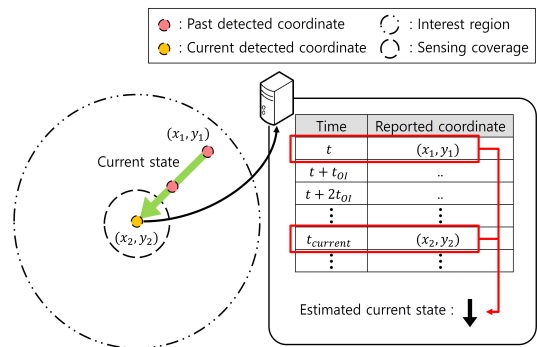


그림 4. 강화 학습을 위한 상태 추정
Fig. 4. State estimation for reinforcement learning

State	1	2	3	4	5	6	7	8	9
$v(km/h)$	$v \leq 3.6$	$v \geq 10$	$v \leq 10$	$v \geq 10$	$v \leq 10$	$v \geq 10$	$v \leq 10$	$v \geq 10$	$v \leq 10$
Direction	(stop)	↑		↓		←		→	
State	-	10	11	12	13	14	15	16	17
$v(km/h)$	-	$v \geq 10$	$v \leq 10$	$v \geq 10$	$v \leq 10$	$v \geq 10$	$v \leq 10$	$v \geq 10$	$v \leq 10$
Direction	-	↗		↖		↙		↘	

그림 5. 강화 학습을 위한 상태 정의
Fig. 5. State definition for reinforcement learning

$$\theta = \cos^{-1} \frac{|x_2 - x_1|}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}} \quad (5)$$

다음 상태는 행동 시행 후 T 시간 동안 구한 현재 상태 정보들을 모두 이용하여 객체의 방향과 속력을 추정한다. T 시간 동안 감지된 센서 노드들의 상태 벡터를 모두 더하여 단위 벡터로 만들어 다음 방향을 추정한다. 방향을 정할 때는 x축을 기준으로 $0 \leq \theta \leq 22.5^\circ$, $22.5^\circ < \theta \leq 67.5^\circ$, $67.5^\circ < \theta \leq 90^\circ$ 로 각도를 나누어 방향을 정한다. 또한 속력은 상태 벡터들의 속력을 평균 내어 추정한다. 추정한 방향과 속력으로 다음 상태를 정의한다.

객체를 감지한 센서는 서버로부터 최대 가치 함수를 갖는 행동을 전달 받아 주변 센서로 유니캐스트 한다. 이 때, 행동을 전달 받게 되는 영역을 Q-Learning의 행동으로 설정하며 이는 그림 6과 같이 18가지로 정의한다. 유니캐스트로 행동을 전달 받은 센서는 즉시 모드2로 전환한다. 영역별로 다양한 행동을 정의한 이유는 각 상태에 대한 가장 효율적인 행동을 선택하기 위함이다.

행동 선택에 영향을 미치는 가치 함수를 계산하여 Q-table을 갱신하기 위해 보상함수가 필요하다. R 은 센서가 행동을 수행한 후 객체를 감지한 센서의 수에 대한 보상을 주는 함수이다. 센서가 객체를 감지하면 행동이 바로 이행되며 그 순간부터 T 시간동안 행동 범위 안에서 객체를 감지한 모드2 상태의 센서 개수와 행동 범위 밖에서 객체를 감지한 모드1인 센서의 개수를 통해 보상 함수가 계산된다. 다음 상태와 마찬가지로 행동을 취하고 정해진 시간 T 동안 데이터를 모은 후 계산하여 Q-table을 갱신한다. 보상 함수는 수식 (6)과 같다.

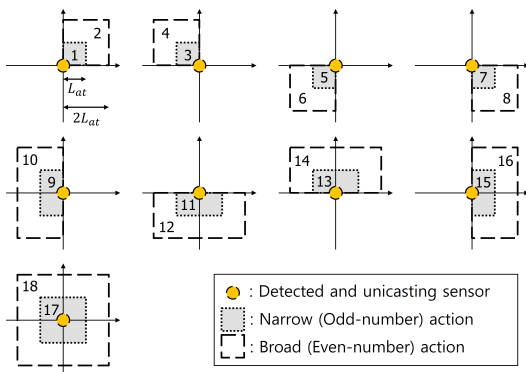


그림 6. 강화 학습을 위한 행동 정의
Fig. 6. Action definition for reinforcement learning

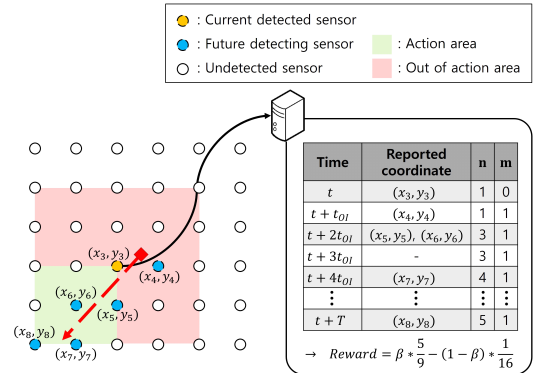


그림 7. 강화 학습을 위한 보상 함수 계산
Fig. 7. Reward calculation for reinforcement learning

$$R = \beta * R_p - (1 - \beta) * R_n \quad (6)$$

수식의 β 는 양의 보상 함수 R_p 와 음의 보상 함수 R_n 의 가중치를 나타낸다. 양의 보상 함수는 행동이 제대로 이행 되었을 때 주는 보상이고, 음의 보상 함수는 행동이 제대로 이행되지 않았을 때 주는 보상이다. 결과적으로 행동이 제대로 이행된다면 보상은 양의 값이지만 행동이 제대로 이행되지 않는다면 보상은 음의 값이 된다. 그림 7과 같이 양의 보상 함수는 행동 영역에서 계산되고 양의 보상 함수는 행동 영역 밖에서 계산된다. 양의 보상 함수 R_p 와 음의 보상 함수 R_n 은 수식 (7),(8)과 같이 정의한다.

$$R_p = \frac{n_{ps}}{N_{ps}} \quad (7)$$

$$R_n = \frac{n_{ng}}{N_{ng}} \quad (8)$$

수식에서 N_{ps} 은 행동 영역의 전체 센서의 수를 의미하며 n_{ps} 은 행동 영역에 속하며 객체를 감지한 센서의 개수이다. N_{ng} 은 행동 영역 밖에 있는 전체 센서의 수를 의미하며 n_{ng} 은 행동 영역 밖에 속하며 객체를 감지한 센서의 개수이다.

2.3 평가 방법

본 절에서는 앞서 제안한 동적 센싱 모드 제어 방법의 성능 평가를 위한 지표로 감지 정확도1과 감지 정확도2, 에너지 소모율을 설명한다. 세 가지 지표 모두 t_{OI} 간격으로 관찰한 결과를 누적하여 계산한다. 또한 정확도 및 에너지 소모율에 대한 기준으로는 객

표 1. 기존 방법과의 비교
Table 1. Comparisons with conventional methods

Type	Comparison	Content	
Fixed Mode	method	Use a predefined duty cycle mode	
	pros	Simple structure without changing the mode	
	cons	A trade-off exists between the detection accuracy and the energy efficiency according to the predefined mode	
	complexity	the simplest structure	
Object Location Estimation Based Minimum Area Selection	type	Dead-reckoning	Kalman filter
	method	Use vector addition for a next position estimation	Use a recursive filter to reduce noise in the next position estimation.
	pros	Relatively simple way to implement	Excellent prediction performance (appropriate for linearity)
	cons	When the movement of the object is nonlinear, prediction error increases, then it causes a poor detection accuracy.	
	complexity	Simplicity	Covariance estimation
Reinforcement Learning Based Dynamic Area Selection	method	Dynamically select the active area by learning through rewards received from the environment	
	pros	Select an appropriate active area considering the movement of the object	
	cons	In online learning, This method requires sufficient time to converge.	
	complexity	Unlike other methods, it requires additionally some simple computations for updating the Q-table	

체의 이동 경로, 센서 배치 상황 등 모든 실험 조건이 동일한 상태에서 모든 센서가 모드2로만 동작하는 경우로 설정한다.

먼저 감지 정확도1은 센서 노드가 객체를 감지하면 즉시 모드2로 전환하고 객체가 감지 반경을 벗어날 때까지 모드2를 유지하는 센서 노드의 동작 특성을 이용한 지표이다. n_{DS} 는 감지한 누적 센서 수이다. 이때 한 센서 노드의 연속된 객체 감지는 최초 감지 시점만 인정하여 1회로 인정한다. 만약 이후 객체가 감지 영역을 벗어나 연속적인 감지 상황이 종료된 후 다시 또 해당 센서의 감지 영역 안으로 객체가 진입한다면 이때는 두 번째 감지상황으로 인정하여 n_{DS} 를 누적한다. n_{M2} 는 n_{DS} 와 계산 방법은 동일하지만, 모든 센서가 모드2로만 동작하는 환경에서의 감지 센서 수이다. 감지 정확도1은 수식 (9)와 같다.

$$\text{Detection Accuracy1} = \frac{n_{DS}}{n_{M2}} \quad (9)$$

감지 정확도2는 감지 정확도1에서 배제한 연속적

인 객체 감지 상황을 독립적으로 인정한 지표이다. m_{DS} 는 매 t_{OI} 마다 센서를 관측하여 객체를 감지한 누적 센서 수이다. m_{M2} 는 모든 센서가 모드2로만 동작하는 환경에서의 객체를 감지한 누적 센서 수이다. 감지 정확도2는 수식 (10)과 같다.

$$\text{Detection Accuracy2} = \frac{m_{DS}}{m_{M2}} \quad (10)$$

에너지 소모율에서는 모든 센서가 모드2로만 동작하는 환경 대비 에너지 소모를 보기 위한 지표이며 수식 (11)과 같다. u_{DS} 는 웨이크업 상태인 누적 센서 수를 나타내며, u_{M2} 는 모든 센서가 모드2로만 동작하는 환경에서 웨이크업 상태인 누적 센서 수를 나타낸다.

$$\text{Energy consumption rate} = \frac{u_{DS}}{u_{M2}} \quad (11)$$

본 논문에서 제안하는 방법은 객체의 이동패턴 상

태에서의 강화학습에 따라 가장 적합한 센싱 노드들의 범위를 현재 상황에 맞게 스스로 학습한다는 것이다. 기존 수학적 방법에 따라 객체의 속도와 방향을 정확히 추정하여 (Dead-reckoning, 칼만 필터), 일정 시간 후 추정된 객체의 위치를 포함하는 가장 작은 범위 내의 센서만을 활성화하는 방법을 적용할 경우 객체의 이동패턴이 예측 가능한 경우에는 좋은 성능을 보이지만, 불규칙적이거나 다양한 패턴이 혼재할 경우 잘못된 위치 추정으로 검출확률을 보장하기 어렵다. 본 논문은 객체를 검출한 노드들의 시 공간 위치를 기반으로 현재 객체의 추정된 속도와 방향을 상태로 정의하고 온라인 실시간 강화학습을 통해 행동 수행 후 보상을 평가함으로써 현재의 이동패턴에 가장 적합한 활성화 센서 영역을 선택한다. 표 1은 기존 Dead-reckoning 및 칼만 필터를 사용한 위치 추정에 의한 최소영역 센서 활성화 방법과 제안된 강화학습 기반 센서 활성화 영역선택 기법의 비교를 보여준다.

III. 모의 실험 및 결과

본 논문이 제안한 무선 센서 네트워크에서 강화 학습 기반의 동적 듀티 사이클 모드 선택 방법의 성능을 확인하기 위한 시뮬레이션을 진행하였다. 시뮬레이션 결과에 따른 최대 가치 함수 값의 변화와 최대 가치 함수를 가지는 행동의 변화를 통해 본 논문이 제안하는 Q-Learning 모델의 수렴을 확인하였다. 앞서 2.3절

표 2. 시뮬레이션 파라미터
Table 2. Simulation parameters

Parameter	Value
Observation time interval t_{OI}	0.1s
Simulation time T_s	2000s
Map size	$100 \times 100m^2$
Sensor deployment interval d_s	3, 6m
The number of sensors N_s	1089, 256
Radius of sensing coverage R_C	2.2m
Mode1 total period mtP_1	5s
Mode1 duty cycle mdc_1	10%
Mode2 holding time T_{hold}	3s
Mode2 duty cycle mdc_2	100%
v_{min}	5km/h
v_{max}	20km/h
Random selected pause time T_{pause}	5, 10, 15, 20s

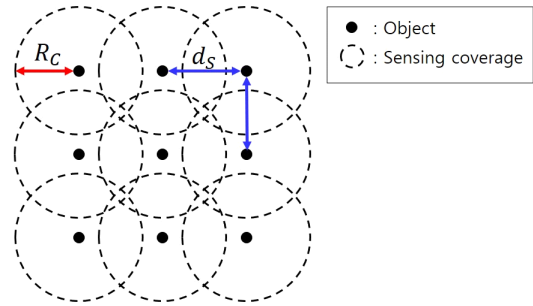


그림 8. 센서 배치
Fig. 8. Sensor deployment

에서 설명한 평가 방법을 이용하여 감지 정확도1, 감지 정확도2, 에너지 소모율을 다양한 센서 네트워크 환경과 다양한 객체 이동 모델 상황에서 확인하였다. 시뮬레이션은 매트랩(Matlab)을 이용하여 진행하였으며 표 2은 시뮬레이션 파라미터를 나타낸다.

센서 네트워크는 $100 \times 100m^2$ 의 정사각형 모양의 환경에 센서를 배치하였다. 센서는 그림 8과 같이 센서 간의 가로와 세로 간격을 d_s 로 하여 격자로 배치하였다. 객체의 이동 모델은 Random waypoint 모델을 변형하여 이용하였다²⁰⁾. 해당 모델은 무작위한 한 점을 목적지로 지정하고 $[v_{min}, v_{max}]$ 범위에서 선택한 속도로 지정한 점까지 이동한다. 객체가 목적지에 도착하면 대기시간 후보군 중 하나를 선택하여 T_{pause} 동안 움직이지 않고 정지한다. 이후 이 과정을 반복한다. 표 3은 시뮬레이션에 사용한 Q-Learning의 파라미터를 나타낸다.

그림 9와 그림 10을 통해 본 논문이 제안하는 Q-Learning 모델의 수렴에 대해서 확인하였다. 그림 9는 특정 상태를 방문한 횟수에 따른 최대 가치 함수 값의 변화를 나타낸다. 그래프의 x축은 특정 상태를 방문한 횟수이며 y축은 특정 상태의 최대 가치 함수이다. 상태를 방문한 횟수가 증가함에 따라 최대 가치

표 3. Q-learning 파라미터
Table 3. Q-learning parameters

Parameter	Value
Length of action area, L_{at}	3m
Q-table updating delay, T	3s
γ	0.9
α	0.2
ϵ_{start}	0.7
ϵ_{end}	0.05
ϵ_{decay}	200
β	0.4

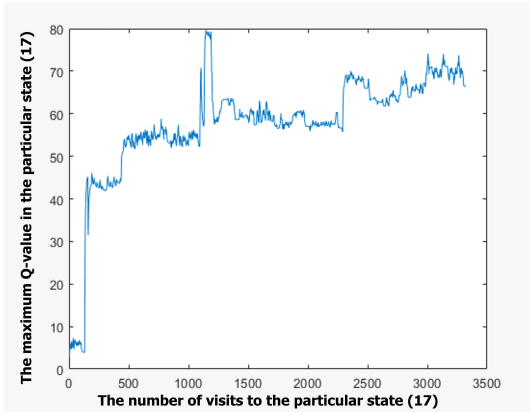


그림 9. 특정 상태를 방문한 횟수에 따른 최대 가치 함수 값의 변화
 Fig. 9. Changes in the value of the maximum Q-value over the number of visits to the particular state

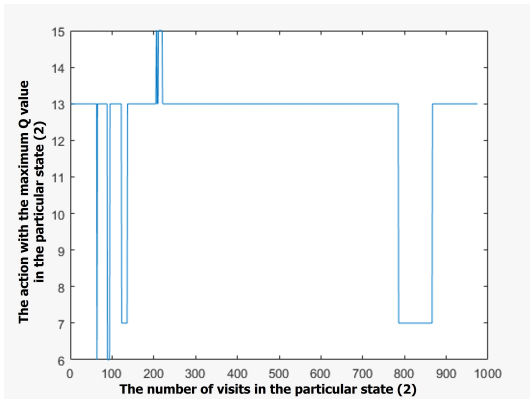


그림 10. 특정 상태를 방문한 횟수에 따른 최대 가치 함수를 갖는 행동의 변화
 Fig. 10. Changes in action with maximum Q-value according to the number of visits to a particular state

함수가 증가하다가 수렴하는 형태를 보이는 것을 확인할 수 있다.

그림 10은 특정 상태에서 최대 가치 함수를 갖는 행동을 나타낸 것이다. 그래프의 x축은 특정 상태를 방문한 횟수이고 y축은 특정 상태의 최대 가치 함수를 갖는 행동을 의미한다. 상태를 방문한 횟수가 증가할수록 최대 가치 함수를 갖는 행동이 수렴하는 것을 확인할 수 있다. 최대 가치 함수를 갖는 행동이 한 개로 수렴하지 않는 이유는 학습의 문제가 아닌 센서 배치 때문에 생기는 결과이다.

본 논문 2.3절에서 설명한 평가 방법을 이용하여 센서의 배치 간격 d_s 가 3m인 경우와 6m인 경우, 그리고 센서를 무작위로 배치한 경우에 대해서 제안한

방식의 성능을 확인하였다. 같은 크기의 네트워크 환경에서 센서의 배치 간격에 따라 d_s 가 3m인 경우에는 배치된 센서의 개수 N_s 가 1089개, d_s 가 6m인 경우에는 배치된 센서의 개수 N_s 가 256개로 개수에 차이가 있다. 센서를 무작위로 배치하는 경우에는 1089개의 센서를 배치하였다. 다양한 센서 네트워크 환경에서 모든 센서가 모드1로 동작하는 센싱 모델과 본 논문이 제안하는 동적 센싱 모드 방법을 비교를 진행하였다. 그림 11-13에서 파란색 지표는 2.3절의 평가 방법에서 설명한 감지 정확도1, 분홍색 지표는 감지 정확도2, 회색 지표는 에너지 소모율을 나타낸다. 그림 11은 d_s 가 3m이고 N_s 가 1089개인 환경에서 모델의 성능을 확인한 것이다. 이 때, 1089개의 센서는 격자로 배치한다. 동적 센싱 모드 제어 방법은 감지 정확도1은 88.9%, 감지 정확도2는 91.3%, 에너지 소모율은 12.3%로 적은 에너지 소모에도 매우 높은 감지 정확도를 보이는 것을 확인하였다. 또한 모든 센서가 모드1로만 동작하는 비교 모델과 비교했을 때, 제안 모델이 감지 정확도1은 53.1%p, 감지 정확도2는 38.5%p 높지만 에너지 소모율은 2.2%p 밖에 증가하지 않은 것을 확인할 수 있다.

그림 12는 d_s 가 6m이고 N_s 가 256개인 환경에서 모델의 성능을 확인한 것이다. 항상 모드1을 유지하는 비교 모델의 경우에는 감지 정확도1은 35.3%, 감지 정확도2는 53.2%, 에너지 소모율은 10.1%로 센서 간의 간격에 구분 없이 비슷한 성능을 보이는 것을 확인할 수 있다. 반면 동적 센싱 모드 제어 방법은 감지 정확도1은 51.2%, 감지 정확도2는

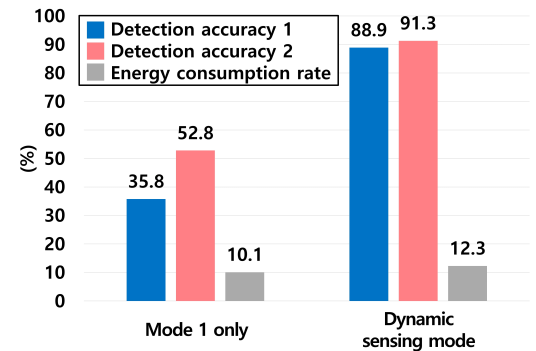


그림 11. $d_s = 3m$, $N_s = 1089$ 에서 모든 센서가 모드1로 동작하는 센서 환경과 동적 센싱 모드 제어 방법 비교
 Fig. 11. Comparison of sensing mode methods with all sensors operating in mode1 and dynamic sensing mode control methods and Dead-reckoning when $d_s = 3m$, $N_s = 1089$

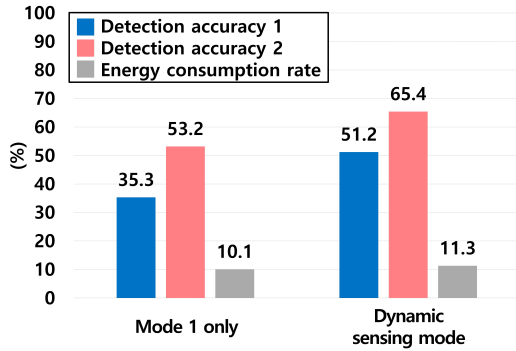


그림 12. $d_s=6m$, $N_s=256$ 에서 모든 센서가 모드1로 동작하는 센서 환경과 동적 센싱 모드 제어 방법 비교
Fig. 12. Comparison of sensing mode methods with all sensors operating in model1 and dynamic sensing mode control methods when $d_s=6m$, $N_s=256$

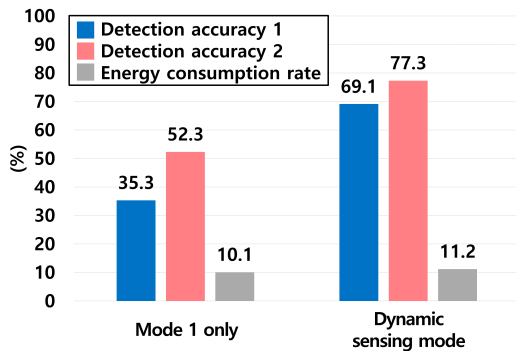


그림 13. $N_s=1089$, 무작위 센서 배치에서의 모든 센서가 모드1로 동작하는 센서 환경과 동적 센싱 모드 제어 방법 비교
Fig. 13. Comparison of sensing mode methods with all sensors operating in model1 and dynamic sensing mode control methods when $N_s=1089$, random sensor deployment

65.4%, 에너지 소모율은 11.3%로 센서 간의 간격을 $3m$ 로 설정했을 때보다 감지 정확도와 에너지 소모율 모두 감소한 것을 확인하였다. 이는 배치 간격을 $3m$ 에서 $6m$ 로 늘리면서 동적 센싱 모드 제어 방법의 행동을 작은 범위와 큰 범위로 구분한 의미가 줄어들었기 때문이다.

그림 13은 센서를 격자 배치가 아닌 무작위로 배치 하였을 때의 결과 그래프로 N_s 가 1089개 일 때 모델의 성능을 확인한 것이다. 센서를 무작위로 배치하였을 때, 동적 센싱 모드의 제어 방법은 감지 정확도1은 69.1%, 감지 정확도2는 77.3%, 에너지 소모율은 11.2%인 것을 확인하였다. 모든 센서가 모드1로 동작하는 모델에 비하여 감지 정확도1은 33.8%p, 감지 정확도2는 25.0%p 높지만 에너지 소모율은 1.1%p 밖에

증가하지 않은 것을 확인할 수 있다. 그림 11과 그림 13에서 동적 센싱 모드 제어 방법의 성능이 비교를 통해 같은 수의 센서를 사용하는 경우 본 논문의 제안 방식이 센서의 배치 환경에는 큰 영향을 받지 않는 것을 알 수 있다.

본 논문 2.3절에서 설명한 평가 방법을 이용하여 세 가지의 객체 이동 모델에 대해서 제안된 방식의 성능을 확인하였다. 센서 네트워크는 센서 간의 간격 d_s 가 $3m$ 이고 센서의 수 N_s 가 1089인 격자 배치로 설정하였다. 기본적인 객체의 이동 방식은 앞서 사용한 변형된 Random waypoint 모델을 사용하였다. 단 실험 환경1의 경우에는 목적지까지 이동하는데 걸리는 시간이 Q-table 갱신 지연 시간인 T 와 같거나 보다 긴 곳에 위치한 목적지를 선택하도록 하였다. 실험 환경2는 전체 센서 필드에서 다음 이동점을 구하고 랜덤하게 선택된 속도와 방향으로 이동하는데, 그림 14와 같이 매 T 시간마다 진행 방향에서 $0\sim 180^\circ$ 범위에서 랜덤하게 방향을 변경하여 이동한다. 객체의 이동 시간이 최단 경로로 이동했을 때 걸리는 시간보다 초과되면 마지막 위치에서 T_{pause} 동안 움직이지 않고 새로운 목적지와 속도를 결정하며 이후 이 과정을 반복한다. 이 방식은 객체가 목적지의 방향으로 이동을 하지만 최종적으로 목적지에 도착하지 않을 수 있다.

실험 환경3은 객체의 목적지가 전체 센서 필드가 아니라 그림 15와 같이 현재 위치에서 두 개의 센서 평균 거리인 $4d_s$ 범위 내로 다음 목적지를 선정하도록 하였다. 목적지의 위치가 가까움에 따라 T_{pause} 를 사용하지 않고 목적지에 도착하게 되면 바로 다음 목적지를 결정해서 이동한다.

본 논문이 제안하는 동적 센싱 모드 방법과 비교 모델로 모든 센서가 모드1로 동작하는 센싱 모델과 Dead-reckoning을 적용한 센싱 모델, 칼만 필터를 적용

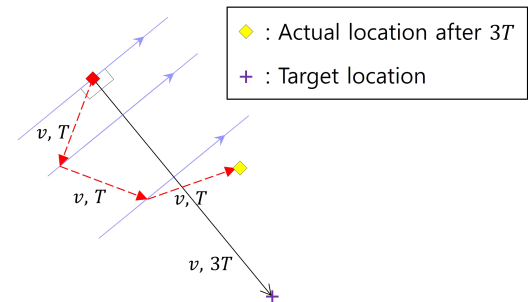


그림 14. 실험 환경2의 객체 이동 방식
Fig. 14. Object movement method in Experimental Environment 2

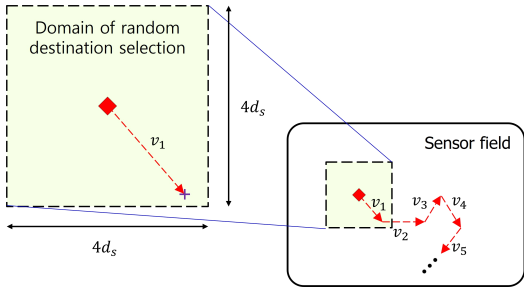


그림 15. 실험 환경3의 객체 이동 방식
Fig. 15. Object movement method in Experimental Environment 3

한 센싱 모델과의 비교를 진행하였다. Dead-reckoning은 기존에 알고 있는 위치로부터 이동방향과 이동거리를 반영하여 현재의 위치를 추정하는 방식이다.^[21] 이는 구성비용이 저렴하고 환경변화에 강인하다는 장점을 가져 위치과악 서비스에 많이 사용된다. 칼만 필터는 추정값과 측정값의 오차에 필터이득을 곱하여 상태천이 모델에 수정함으로써 추정오차를 최소화하는 처리과정이다.^[22] 객체의 위치를 (x_k, y_k) , x축 방향으로의 속력을 v_k^x , y축 방향으로의 속력을 v_k^y 로 표현할 때, 객체의 위치와 축에 따른 속력을 나타내는 상태 벡터 X_k 와 방정식은 각각 수식 (12)-(13)으로 정의된다.^[23]

$$X_k = [x_k, y_k, v_k^x, v_k^y]^T \quad (12)$$

$$X_k = A_k X_{k-1} \quad (13)$$

수식 (13)의 A_k 는 상태 천이 행렬이며 이는 수식 (14)와 같이 정의한다.

$$A_k = \begin{bmatrix} 1 & 0 & t_{OI} & 0 \\ 0 & 1 & 0 & t_{OI} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (14)$$

본 논문에서 사용한 오차 공분산 P_k 의 초기값은 수식 (15)와 같이 정의되며 수식 (16)를 통해 갱신된다.

$$P_0 = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix} \quad (15)$$

$$P_k = A_k P_{k-1} A_k^T + Q_k \quad (16)$$

본 논문에서 칼만 필터 구현을 위해 사용한 시스템 잡음 공분산 행렬 Q_k , 출력 행렬 H_k , 측정 잡음 공분산 행렬 R_k 을 각각 수식 (17)-(19)로 정의했을 때 칼만 이득 K_k 는 수식 (20)을 통해 정의된다.

$$Q_k = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \quad (17)$$

$$H_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (18)$$

$$R_k = \begin{bmatrix} d_s^2 & 0 \\ 0 & \left(\frac{d_s}{m d c_1 \times m t p_1} \right)^2 \end{bmatrix} \quad (19)$$

$$K_k = P_k H_k^T (H_k P_k H_k^T + R_k)^{-1} \quad (20)$$

최종적으로 측정값 z_k 으로부터 상태 벡터와 오차 공분산을 추정할 수 있다.

$$\hat{X}_k = X_k + K_k (z_k - H_k X_k) \quad (21)$$

$$\hat{P}_k = P_k - K_k H_k P_k \quad (22)$$

Dead-reckoning을 적용한 센싱 모델과 칼만 필터를 적용한 센싱 모델의 경우에는 계산된 추정점으로부터 t_{OI} 시간 후 이동한 지점을 예측하고 이를 포함하는 가장 크기가 작은 모드를 선택한다. 그림 16-18에서 파란색 지표는 2.3절의 평가 방법에서 설명한 감지 정확도1, 분홍색 지표는 감지 정확도2, 회색 지표는 에너지 소모율을 나타낸다.

그림 16은 실험환경1에서 모델의 성능을 확인한 것이다. 모든 센서가 모드1로만 동작하는 모델은 감지 정확도1은 36.2%, 감지 정확도2는 53.6%, 에너지 소모율은 10.1%로 에너지 소모율과 감지 정확도 모두 가장 낮은 결과를 보였다. 동적 센싱 모드 제어 방법의 감지 정확도1은 92.2%, 감지 정확도2는 94.1%, 에너지 소모율은 12.1%인 것을 확인할 수 있다. Dead-reckoning을 적용한 센싱 모델의 감지 정확도1은 95.8%, 감지 정확도2는 97.9%, 에너지 소모율은

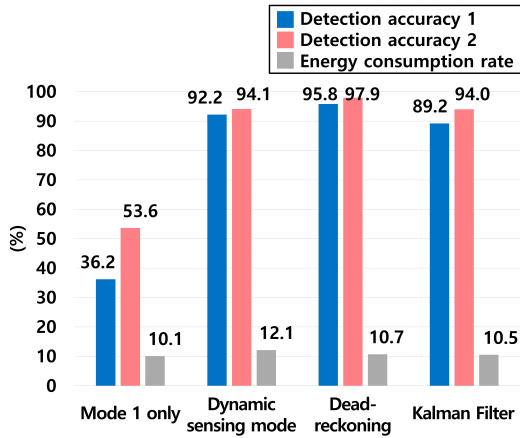


그림 16. 실험환경 1에서 성능 비교
Fig. 16. Performance comparison in experiment environment 1

10.7%인 것을 확인할 수 있다. 칼만 필터를 적용한 센싱 모델의 감지 정확도1은 89.2%, 감지 정확도2는 94.0%, 에너지 소모율은 10.5%로 동적 센싱 모드 제어 방법, Dead-reckoning과 칼만 필터를 적용한 센싱 모델 모두 매우 높은 성능을 보이는 것을 확인할 수 있다.

그림 17은 실험 환경2에서 모델의 성능을 확인한 것이다. 모든 센서가 모드1로 동작하는 모델은 감지 정확도1은 37.1%, 감지 정확도2는 56.5%, 에너지 소모율은 10.1%로 에너지 소모율과 감지 정확도 모두 가장 낮은 성능을 보이는 것을 확인하였다. 동적 센싱 모드 제어 방법의 감지 정확도1은 92.1%, 감지 정확도2는 95.0%, 에너지 소모율은 12.4%인 것을 확인할 수 있다. Dead-reckoning을 적용한 센싱 모델의 감지

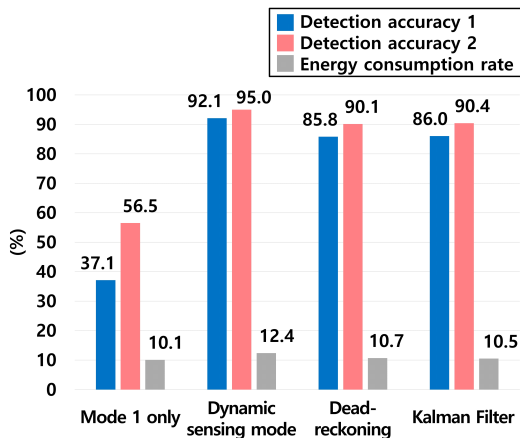


그림 17. 실험환경 2에서 성능 비교
Fig. 17. Performance comparison in experiment environment 2

정확도1은 85.8%, 감지 정확도2는 90.1%, 에너지 소모율은 10.7%인 것을 확인할 수 있다. 칼만 필터를 적용한 센싱 모델의 감지 정확도1은 86.0%, 감지 정확도2는 90.4%, 에너지 소모율은 10.5%인 것을 확인할 수 있다. 이를 통해 실험 환경2와 같이 객체의 움직임이 자주 변하는 환경에서는 Dead-reckoning과 칼만 필터를 적용한 센싱 모델에 비해 본 논문이 제안하는 방식이 약 6%이상 높은 성능을 보이는 것을 확인할 수 있다.

그림 18은 실험환경 3에서 모델의 성능을 확인한 것이다. 모든 센서가 모드1로만 동작하는 모델은 감지 정확도1은 30.4%, 감지 정확도2는 24.0%, 에너지 소모율은 10.0%로 에너지 소모율과 감지 정확도 모두 가장 낮은 성능을 보이는 것을 확인하였다. 동적 센싱 모드 제어 방법의 감지 정확도1은 96.6%, 감지 정확도2는 95.4%, 에너지 소모율은 12.8%인 것을 확인할 수 있다. Dead-reckoning을 적용한 센싱 모델의 감지 정확도1은 66.1%, 감지 정확도2는 62.1%, 에너지 소모율은 10.6%인 것을 확인할 수 있다. 칼만 필터를 적용한 센싱 모델의 감지 정확도1은 84.5%, 감지 정확도2는 79.9%, 에너지 소모율은 10.7%인 것을 확인할 수 있다. 이를 통해 동적 센싱 모드 제어 방법은 여전히 매우 높은 성능을 보이지만 Dead-reckoning을 적용한 센싱 모델과 칼만 필터를 적용한 센싱 모델은 낮은 성능을 보이는 것을 확인하였다.

Dead-reckoning과 칼만 필터를 적용한 센싱 모델의 경우에는 실험 환경1과 같이 객체의 이동에 대한 예측이 쉬운 환경에서는 높은 성능을 보이는 것을 확인할 수 있다. 반면 실험 환경2와 같이 거시적인 관점

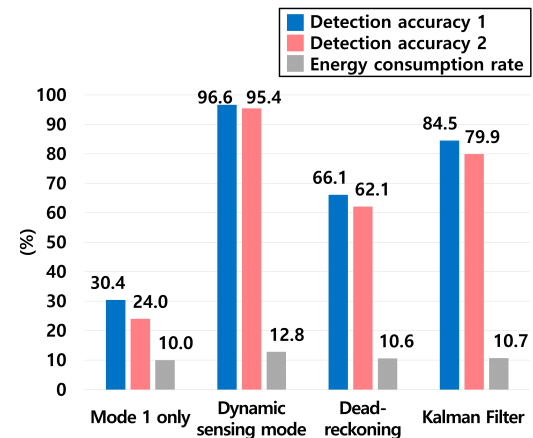


그림 18. 실험환경 3에서 성능 비교
Fig. 18. Performance comparison in experiment environment 3

에서는 목적지의 방향으로 이동하지만 미시적인 관점에서 움직임에 대한 예측이 어려운 경우에는 다소 성능이 떨어지고 실험 환경3과 같이 객체의 이동 속도와 방향이 자주 바뀌는 환경에서는 낮은 성능을 보이는 것을 확인할 수 있다. 이에 반해 본 논문이 제안하는 방식인 동적 센싱 모드 제어 방법은 갑작스러운 객체의 움직임 변화에 의해 객체의 이동에 대한 예측이 어려운 환경에도 높은 성능을 보이는 것을 확인할 수 있다. 이는 Q-Learning 학습 시 이러한 객체의 움직임에 대한 불확실성에 대해서도 학습하기 때문이다. 계산 비용의 관점에서 본 논문이 제안하는 방식과 Dead-reckoning, 칼만필터를 적용한 방식 모두 T 시간마다 서버로 센싱 결과를 보고하고 이를 통해 객체의 이동 속도를 계산하는 것, 그리고 새로운 모드 변경 정보를 다시 노드로 전송하는데 필요한 비용은 같다. 하지만 Dead-reckoning을 적용한 센싱 모델에 비해 칼만 필터를 적용한 센싱 모델의 경우에는 칼만 이득을 계산하기 위해 추가적으로 공분산 계산에 대한 비용이 발생하게 된다. 또한 제안된 방법의 경우에는 Q-Learning 학습에 필요한 보상 계산과 Q-table 갱신에 대한 비용이 추가적으로 발생하게 된다.

IV. 결 론

객체 추적을 위한 무선 센서 네트워크에서 객체 추적 정확성과 센서의 에너지 효율성은 서로 트레이드오프 관계이다. 네트워크의 수명을 최대화하는 동시에 객체 추적의 목적을 잃지 않기 위해서는 두 요소를 모두 고려한 최적의 센서 동작을 찾는 것이 매우 중요하다. 일반적인 상황에서 객체의 이동성을 예측하고 최적의 동작을 실시간으로 계산하는 것은 어려운 일이다. 따라서 우리는 강화 학습의 기법 중 하나인 Q-Learning을 이용하여 객체의 이동성을 고려한 최적의 듀티 사이클을 산출한다. Q-Learning은 주어진 환경과 상태로부터 비지도 학습을 통해 최적의 결과를 산출하는 방식이다. 따라서 본 논문은 Q-Learning 모델을 학습시키기 위한 네트워크 환경, 에이전트의 상태와 행동에 대해서 제시하였으며 모의실험을 통해 제안된 방식의 성능을 확인하였다. 제안된 Q-Learning 모델이 시뮬레이션이 경과함에 따라 수렴하는 것을 확인하였고 감지 정확도1, 감지 정확도2, 에너지 소모율에 대한 성능 평가를 통해 제안 방식이 매우 높은 성능을 보이는 것을 확인하였다. 본 논문이 제안하는 방식은 실내, 실외 구분 짓지 않고 다양한 환경에서 화재, 가스, 사람, 동물 등 다양한 객체 추적

에 대해 적용이 가능하다. 이를 통해 서비스 시스템 운영비용은 감소시키며 객체 추적의 정확성을 보장할 수 있는 무선 센서 네트워크 시스템을 제공할 수 있을 것으로 기대된다.

References

- [1] S. M. Jung, "Multi-Chain based routing protocol for efficient energy consumption in wireless sensor networks," *J. Digital Contents Soc.*, vol. 21, no. 6, pp. 1181-1189, 2020.
- [2] W. K. Yun and S. J. Yoo, "Q-Learning-Based data-aggregation-aware energy-efficient routing protocol for wireless sensor networks," *IEEE Access*, vol. 9, pp. 10737-10750, 2021.
- [3] D. H. Lee, K. W. Cho, C. G. Kang, and C. H. Oh, "Link cost based routing protocol for improving energy efficiency in wireless sensor networks," *J. KIICE*, vol. 23, no. 5, pp. 574-580, 2019.
- [4] D. G. Hwang, H. Park, J. H. Park, D. O. Seong, and J. S. Yoo, "Moving object tracking scheme based on polynomial regression prediction in sparse sensor networks," *The J. Korea Contents Assoc.*, vol. 12, no. 3, pp. 44-54, 2012.
- [5] J. Luo, Z. Zhang, C. Liu, and H. Luo, "Reliable and cooperative target tracking based on WSN and WiFi in indoor wireless networks," in *IEEE Access*, vol. 6, pp. 2169-3536, 2018.
- [6] S. W. Jang, J. S. Hahn, and R. Ha, "An energy-efficient edge detection method for continuous object tracking in wireless sensor networks," *J. KISS: Info. Netw.*, vol. 36, no. 6, pp. 514-527, 2009.
- [7] B. Jiang, B. Ravindran, and H. J. Cho, "Probability-based prediction and sleep scheduling for energy-efficient target tracking in sensor networks," *IEEE Trans. Mob. Comput.*, vol. 12, no. 4, pp. 735-747, 2012.
- [8] J. H. Jeon, "Efficient mechanism for receiver and sink node in wireless sensor networks," *J. Korean Assoc. Comput. Edu.*, vol. 23, no. 3, pp. 65-70, 2020.

- [9] W. K. Yun, Y. J. Song, J. S. Moon, S. J. Jang, and S. J. Yoo, "Deep learning based temperature sensor data and wildfire propagation prediction in duty cycled wireless sensor network," *J. KICS*, vol. 44, no. 6, pp. 1092-1104, 2019.
- [10] S. W. Lee, H. S. Yoo, and D. K. Kim, "Residual energy-aware duty-cycle scheduling scheme in energy harvesting wireless sensor networks," *J. KICS*, vol. 39, no. 10, pp. 691-699, 2014.
- [11] H. K. Lee, "Design of a multipath routing protocol for energy-efficiency and low-delay in duty-cycled wireless sensor networks," *J. IKEEE*, vol. 24, no. 1, pp. 326-332, 2020.
- [12] M. H. Kim and J. S. Hong, "An energy-efficient and practical duty-cycling mechanism on building automation and home sensor networks," *J. Korean Soc. Internet Info.*, vol. 13, no. 3, pp. 23-30, 2012.
- [13] K. S. Park, J. M. Park, W. K. Yun, and S. J. Yoo, "DQN reinforcement learning: The robot's optimum path navigation in dynamic environments for smart factory," *J. KICS*, vol. 44, no. 12, pp. 2269-2279, 2019.
- [14] J. H. Lee, G. W. Choi, and J. W. So, "Q-learning based downlink MIMO-NOMA scheme for vehicle networks," *J. KICS*, vol. 44, no. 8, pp. 1493-1503, 2019.
- [15] J. H. Lee and D. S. Eom, "The energy performance analysis between synchronous and asynchronous duty cycle based mac protocols in wireless sensor networks," *J. Info. Technol. Serv.*, vol. 10, no. 3, pp. 237-250, 2011.
- [16] E. G. Ham and C. B. Kim, "Model implementation of reinforcement learning for trading prediction using deep Q network," *J. KIIT*, vol. 17, no. 4, pp. 1-8, 2019.
- [17] H. W. Kim and W. C. Lee, "Real-time path planning for mobile robots using Q-Learning," *J. IKEEE*, vol. 24, no. 4, pp. 991-997, 2020.
- [18] S. H. Choi and S. J. Yoo, "Q-Learning based optimal escape route decision in a disaster environment," *J. KICS*, vol. 46, no. 4, pp. 638-650, 2021.
- [19] B. K. Lee and Y. A. Ahn, "An action select reinforcement learning model in multi agent environment," *J. Knowledge Info. Technol. and Syst.*, vol. 10, no. 4, pp. 455-463, 2015.
- [20] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, "Stochastic properties of the random waypoint mobility model," *Wirel. Netw.*, vol. 10, pp. 555-567, 2004.
- [21] H. Yi and J. S. Kim, "A dead reckoning calibration approach using activity recognition for indoor pedestrian localization," *J. IEIE*, vol. 57, no. 1, pp. 70-71, 2020.
- [22] T. S. Jin, "Estimation of people tracking by kalman filter based observations from laser range sensor," *J. The Korean Soc. Ind. Convergence*, vol. 22, no. 3, pp. 265-272, 2019.
- [23] H. D. Cho, "Target localization method based on extended kalman filter using multipath time difference of arrival," *Korea Academy Ind. Cooperation Soc.*, vol. 22, no. 6, pp. 251-257, 2021.

윤 성 호 (Seong-Ho Youn)



2016년 2월~현재 : 인하대학교
정보통신공학과 학사
<관심분야> 네트워크, 통신 인프라, 머신러닝, 소프트웨어 개발

윤 완 규 (Wan-Kyu Yun)



2019년 2월 : 인하대학교 정보통신공학과 학사 졸업
2019년 2월~현재 : 인하대학교 정보통신공학과 석사
<관심분야> 이동통신, IoT, 무선 센서 네트워크, 머신러닝

최 수 연 (Sue-Yeon Choi)



2018년 2월~현재 : 인하대학교 정보통신공학과 학사
<관심분야> 디지털 시스템 설계, 아날로그 회로 설계, 디지털 통신

최 승 희 (Seung-Hee Choi)



2020년 2월 : 인하대학교 정보통신공학과 학사 졸업
2020년 2월~현재 : 인하대학교 전기컴퓨터공학과 석사
<관심분야> 이동통신, IoT, 무선 센서 네트워크, 머신러닝, 강화학습

김 소 명 (So-Myeong Kim)



2018년 2월~현재 : 인하대학교 정보통신공학과 학사
<관심분야> 네트워크, 이동통신, 통신 인프라, 프로토콜

유 상 조 (Sang-Jo Yoo)



1988년 2월 : 한양대학교 전자통신학과 (공학사)
1990년 2월 : 한국과학기술원 전기및전자공학과 (공학석사)
2000년 8월 : 한국과학기술원 전자전산학과 (공학박사)
1990년 3월~2001년 2월 : KT 연구 개발 본부

1990년 3월~2000년 11월 : NIST(미국 표준기술연구원) 초빙연구원

2001년 3월~현재 : 인하대학교 정보통신공학과 교수
<관심분야> 무선 네트워킹 프로토콜, Cross-layer 프로토콜 설계, Cognitive Radio Network, 무선센서네트워크, 미래인터넷

[ORCID:0000-0003-1533-0814]