



DOI 10.28925/2663-4023.2025.29.907

УДК 004.896

Ганенко Людмила Дмитрівна

аспірантка кафедри комп'ютерної інженерії,

Державний університет інформаційно-комунікаційних технологій, Київ, Україна

ORCID ID: 0000-0003-2219-8196

hanenkoliudmyla@gmail.com**Жебка Вікторія Вікторівна**

доктор технічних наук, професор,

завідувач кафедри технологій цифрового розвитку,

Державний університет інформаційно-комунікаційних технологій, Київ, Україна

ORCID ID: 0000-0003-4051-1190

viktoria_zhebka@ukr.net

МОДЕЛЬ СОЦІАЛЬНО-АДАПТИВНОЇ НАВІГАЦІЇ МОБІЛЬНОГО РОБОТА З ВИКОРИСТАННЯМ МЕТОДІВ НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Анотація. Класичні алгоритми планування траєкторії, незважаючи на свою ефективність у статичних середовищах, демонструють суттєві обмеження при інтеграції в динамічне соціальне середовище. Основний недолік полягає у нездатності в режимі реального часу інтерпретувати рух людини, що призводить до непередбачуваних і потенційно небезпечних маневрів. У відповідь на ці обмеження, набули значної популярності методи навчання з підкріпленням (Reinforcement Learning, RL). Методи RL дозволяють автономному мобільному роботу самостійно формувати оптимальну стратегію поведінки шляхом безпосередньої взаємодії з оточенням та отримання зворотного зв'язку у вигляді винагород чи штрафів. Дане дослідження фокусується на методах навчання з підкріпленням та соціальних моделях поведінки з метою розробки безпечної, ефективною та соціально-адаптивною навігації для автономного мобільного робота. В цій роботі представлено модель, ключовою особливістю якої є комплексний підхід до формування поведінки агента. Запропонована модель враховує не лише базові завдання, досягнення цілі та уникнення фізичних перешкод, але й важливі аспекти соціальної взаємодії. Наукова новизна роботи полягає у розробці багатокомпонентної функції винагороди, яка інтегрує винагороди за досягнення цільової точки, уникнення зіткнень з динамічними й статичними об'єктами, а також цілеспрямовано заохочує агента дотримуватись соціально прийнятних норм. Таким чином, робот навчається не лише уникати людей, а й робити це у спосіб, який є інтуїтивно зрозумілим та комфортним для них. Кінцевою метою дослідження є створення навігаційного агента, який є не лише безпечним, але й соціально інтелектуальним. Це є кроком на шляху до повноцінної інтеграції автономних роботизованих систем у повсякденне людське середовище, оскільки успішне співіснування вимагає не лише фізичної безпеки, але й психологічного комфорту та інтуїтивної зрозумілості поведінки робота.

Ключові слова: інформаційні технології; моделювання; методи машинного навчання; методи навчання з підкріпленням; автономні мобільні роботи; навігація мобільних роботів.

ВСТУП

Впровадження автономних мобільних роботів у повсякденне життя людини — від логістичних складів та виробничих цехів до медичних закладів та громадських просторів — стало актуальним завданням сучасної робототехніки. Ефективна та безпечна навігація є фундаментальною умовою для успішної інтеграції людини та робота. Проте, якщо навігація у статичних середовищах є значною мірою вирішеною



задачею, то присутність людей у середовищі створює значні виклики для мобільного робота. Люди є непередбачуваними динамічними перешкодами, які вимагають від робота не лише здатності уникати зіткнень, але й розробки поведінки, яка є соціально прийнятною та не викликає дискомфорту в оточуючих.

Класичні алгоритми планування шляху є недостатньо гнучкими для адаптації до непередбачуваних рухів людей у реальному часі. У відповідь на ці обмеження, методи навчання з підкріпленням (Reinforcement Learning, RL) набули значної популярності завдяки своїй здатності навчатися складним стратегіям поведінки безпосередньо із взаємодії із середовищем.

Уникнення зіткнень є недостатнім для комфортної взаємодії автономного мобільного робота з людиною, тому наше дослідження виходить за рамки традиційного підходу уникнення перешкод, фокусуючись на створенні соціально-адаптивної моделі.

Постановка проблеми. Класичні алгоритми планування шляху розглядають людей як динамічні перешкоди і не враховують правил взаємодії з людиною. Такий підхід призводить до того, що поведінка робота, хоч і є безконфліктною з точки зору уникнення зіткнень, часто виявляється неефективною, непередбачуваною та психологічно дискомфортною для оточуючих.

Хоча методи навчання з підкріпленням (RL) демонструють значний потенціал для вирішення цієї задачі, більшість існуючих досліджень фокусуються на окремих аспектах соціальної поведінки, наприклад, лише на дотриманні проксемічних зон. Залишається невирішеною проблема створення цілісної моделі, яка б одночасно враховувала та збалансовувала ключові соціальні критерії.

Аналіз останніх досліджень і публікацій. Розробка навігаційних систем мобільних роботів, які працюють у динамічних середовищах із присутністю людей, є складним завданням через необхідність поєднання ефективності планування шляху та соціальної прийнятності руху. Навігація з урахуванням присутності людини є перетином досліджень із планування руху робота та взаємодії людини і робота (human robot interaction, HRI) [1]. Ефективна та безпечна навігація вимагає від робота здатності розуміти людську поведінку та діяти за правилами співпраці.

Під час руху в динамічному середовищі робот повинен швидко реагувати на рухомі перешкоди, враховуючи свої кінематичні та динамічні обмеження. Автори [2] провели дослідження щодо адаптації існуючих традиційних алгоритмів планування шляхів (D*(Dynamic A*), D* Lite, RRT) для роботи в динамічних середовищах. В роботі [3] запропоновано алгоритм, який інтегрує вдосконалений алгоритм A* з методом динамічного вікна (DWA), тим самим усуває недоліки алгоритму A* і запобігає потраплянню мобільного робота в локальний оптимум. Зазначені методи генерують траєкторії без зіткнень, але вони не завжди враховують соціальні обмеження, такі як проксеміка, людський комфорт, природність. З цією метою в [4] представлено варіант динамічного вікна (DWA), який розширює функцію оптимізації для задач навігації соціального робота. У роботі [5] застосовано крок прогнозування до алгоритму DWA та використано адаптивну нейронечітку систему виведення для коригування вагових коефіцієнтів функції оптимізації DWA.

Інший підхід до реалізації соціальної навігації полягає у використанні карт витрат на основі теорії проксеміки. Такий підхід передбачає створення динамічного шару проксемічної карти витрат та додавання гаусових витрат навколо виявленої динамічної або статичної перешкоди. Таким чином, робот переміщується серед людей без зіткнень, поважаючи їхні особисті зони [6]. Однак, обчислювальні вимоги такого підходу в рази зростають, коли збільшується кількість людей або розмір карти середовища. Більше того,



якщо карти витрат однієї або кількох динамічних чи статичних перешкод навколо робота блокують можливі шляхи його руху, робот не може рухатись, навіть, якщо існуватиме доступний простір для дії.

Модель соціальних сил (Social Force Model, SFM) є популярним підходом до опису руху пішоходів. Запропоновано багато розширень цієї моделі для моделювання різних соціальних взаємодій. SFM використовуються не лише для прогнозування майбутньої поведінки людини, але й як локальний планувальник для роботів, який генерує соціально прийнятні траєкторії. Робота [7] присвячена розробці системи соціальної навігації для мобільних сервісних роботів із використанням SFM. Забезпечення комфортного співіснування робота і людини здійснюється шляхом врахування соціальних сигналів та передбачення потенційних зіткнень.

У контексті соціальної навігації дослідники широко використовують методи навчання з підкріпленням (Reinforcement Learning, RL). Застосування алгоритмів навчання з підкріпленням до задачі локальної навігації продемонструвало перспективні результати [8]. У працях [9], [10] RL було використано для навігації в середовищах із перешкодами, але без урахування соціальних аспектів. У дослідженні [11] враховано проксемічні зони у функції винагороди для зменшення порушень соціальних норм. Автори продемонстрували, що соціально прийнятна поведінка мобільного робота може бути сформована в процесі навчання через штрафи за порушення норм, без необхідності точної імітації людських дій.

Однією з ключових переваг підходу, заснованого на навчанні, є врахування аспектів соціальної взаємодії та реагування на зворотний зв'язок від людини. Отже, підходи, засновані на навчанні, мають потенціал для значного покращення галузі соціально свідомої навігації роботів, дозволяючи роботам орієнтуватися та взаємодіяти з людьми більш природним та інтуїтивним чином.

Незважаючи на значний прогрес у дослідженні соціальної взаємодії мобільних роботів, залишається невирішеною проблема моделювання взаємодії з людиною. Функції винагороди часто фокусуються на одному аспекті соціальної поведінки. Недостатньо досліджено синергетичний ефект від поєднання кількох соціальних факторів.

Мета статті. Метою даного дослідження є розробка моделі соціально-адаптивної навігації мобільного робота на основі методів навчання з підкріпленням, яка забезпечує ефективне та соціально прийнятне переміщення робота у динамічному соціальному середовищі.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

В робототехніці навігація визначається як цілеспрямована діяльність робота щодо зміни його просторового положення для досягнення поставленої мети в певному середовищі [12]. Завдання навігації мобільних роботів у динамічних середовищах вимагає здатності ефективно та безпечно пересуватися поблизу людей. Проблема полягає у формалізації ледь помітних соціальних норм, яких дотримуються люди, оскільки їх складно виразити у вигляді чітких правил чи метрик.

Необхідність моделювання поведінки мобільного робота в середовищі із присутністю людей призвела до формування окремого напрямку досліджень, що характеризується власним термінологічним апаратом. У науковій літературі, присвяченій робототехніці, дана проблематика отримала кілька термінологічних визначень, серед яких найбільш поширеними є «соціальна навігація» (social navigation),



«соціально усвідомлена навігація» (socially aware navigation), «соціально-адаптивна навігація» (socially adaptive navigation).

Аналіз цих термінів виявляє суттєві семантичні відмінності. Термін «соціальна навігація» робить акцент на інтеграції соціальних норм та протоколів, що є характерними для людської взаємодії. Проте, такий підхід несе в собі ризик термінологічної некоректності, оскільки може створювати хибне уявлення про ідентичність соціальних можливостей робота та людини.

У сучасних дослідженнях з автономної навігації мобільних роботів терміни «соціально усвідомлена навігація» [13] та «соціально-адаптивна навігація» [14] **Помилка! Джерело посилання не знайдено.** часто вживаються як синонімічні. Обидва поняття позначають здатність робота враховувати соціальні контексти, правила людської поведінки та психологічний комфорт людини під час планування руху, водночас не вимагаючи від робота повної імітації людської поведінки. Терміни охоплюють один і той самий набір вимог: забезпечення передбачуваної та безпечної поведінки робота в соціальному середовищі.

У нашій роботі ми надаємо перевагу терміну «соціально-адаптивна навігація», оскільки він відображає ключовий фокус нашого підходу — динамічну адаптацію навігаційної поведінки робота до змінних сценаріїв взаємодії з людьми. Обраний термін підкреслює здатність системи не просто усвідомлювати соціальні правила, а й активно пристосовуватися до них. Такий підхід є релевантним для досліджень в галузі адаптивного планування руху мобільного робота в середовищі із присутністю людей.

Роботизований агент може вважатися соціально-адаптивним, якщо він відповідає наступним критеріям: агент здатен виявляти людей у середовищі, ідентифікувати їх як особливий клас об'єктів та розглядати їхню безпеку як найвищий пріоритет своєї системи прийняття рішень; у ситуаціях навігаційного конфлікту агент здійснює оцінку контексту та обирає стратегію поведінки, що відповідає соціальним нормам; поведінка агента спроектована для мінімізації порушень особистого простору та створення комфорту для людей [15].

Одним із ключових параметрів, які визначають соціально-адаптивну навігацію, є дистанція, якої робот повинен дотримуватись відносно людини. У цьому контексті, теоретичною основою для моделювання такої поведінки слугує проксемика.

Проксемика — це галузь психологічних досліджень, яка вивчає використання простору у соціальній взаємодії. Концепцію віртуального особистого простору навколо людини запропонував Едвард Т. Холл. Він виділив проксемічні зони, які люди обирають для взаємодії: інтимну, персональну, соціальну та публічну [16].

Інтимна зона — це особистий простір у радіусі до 50 см, присутність у якому дозволено лише найближчим людям. Необґрунтоване вторгнення в інтимну зону є максимально дискомфортним для людини. Особиста зона (0,5–1,2 м) — це простір для комфортної взаємодії з друзями та знайомими. Соціальна зона (1,2–3,5 м) призначена для формального спілкування, без фізичного контакту. Публічна зона (понад 3,5 м) — це дистанція, якої дотримуються під час публічних виступів та перебування серед незнайомих.

Розмір проксемічних зон може змінюватись залежно від ситуації. У таких місцях, як ліфт або громадський транспорт, люди стоять дуже близько, порушуючи особисті та інтимні зони один одного. Однак це не вважається вторгненням, оскільки усвідомлюється, що ситуація не дозволяє дотримуватись відповідної дистанції. Таким чином, іноді радіус публічної зони може тимчасово дорівнювати радіусу особистої або інтимної зони. Визначення проксемічних зон залежить як від окремої людини, так і від соціального та культурного контексту [17].



Варто зазначити, що люди більш вимогливі щодо свого фронтального простору, вважаючи рух назустріч у фронтальній зоні найбільш незручним [18].

Таким чином, мобільний робот повинен не лише уникати фізичного контакту з людиною, але й планувати свою траєкторію так, щоб мінімізувати соціальний дискомфорт [19].

Використання методів RL в навігації мобільних роботів. Завдання навігації є однією з ключових проблем у галузі мобільної робототехніки. Серед сучасних підходів до її розв'язання методи навчання з підкріпленням привертають значну увагу дослідників завдяки високій здатності до узагальнення, а також можливості формувати ефективну навігаційну поведінку на основі взаємодії з навколишнім середовищем. Серед переваг RL можна зазначити відсутність карт, високу здатність до навчання та низьку залежність від точності датчиків [20].

Навчання з підкріпленням використовується для вирішення задач послідовного прийняття рішень, де агент взаємодіє із середовищем для досягнення певної мети шляхом максимізації кумулятивної винагороди. Ці задачі формуються як процес прийняття рішень Маркова (MDP), який характеризується кортежем

$$(S, A, P, R, \gamma),$$

де

S — простір станів, що представляє всі можливі стани середовища. Стан $s_t \in S$ відображає всю відповідну інформацію про навколишнє середовище у заданий момент часу t .

A — простір дій, що представляє всі можливі дії, які може виконати агент. Дія $a_t \in A$ — це дія, застосована агентом у момент часу t .

$P(s' | s, a)$ — ймовірність переходу зі стану $s_t \in S$ до наступного стану $s_{t+1} \in S$ в результаті виконання дії $a_t \in A$. Ця функція фіксує динаміку середовища та може бути детермінованою або стохастичною.

$R(s, a)$ — функція винагороди, яка призначає скалярну винагороду кожній парі стан-дія (s, a) . Винагорода кількісно визначає безпосередню бажаність виконання дії у стані s . Вона може відображати заохочення за близькість до мети або штрафи за небажані дії, такі як зіткнення.

γ — коефіцієнт дисконтування $\gamma \in [0, 1)$, який визначає відносну важливість негайних та майбутніх винагород. Менше значення γ робить більший акцент на негайних винагородах, тоді як значення ближче до 1 заохочує довгострокове планування.

У контексті планування шляху автономного мобільного робота метою RL є обчислення політики, яка дозволяє роботу рухатись з початкової позиції до цільової, уникаючи перешкоди та мінімізуючи витрати на шлях. Формулювання цієї задачі за допомогою MDP описується наступним чином.

Простір станів (S): кожен стан $s \in S$ відображає поточний стан робота та навколишнє середовище. Наприклад, s може включати швидкість, положення (x, y) , орієнтацію θ робота та показники датчиків.

Простір дій (A): дія $a \in A$ — вхідний керуючий сигнал, який подається роботу. Наприклад, a може визначати дискретні команди руху або безперервні керуючі сигнали, такі як лінійна швидкість v та кутова швидкість ω .

Перехід стану ($P(s' | s, a)$) моделює динаміку руху робота та його взаємодію з навколишнім середовищем. Для робота з кінематикою диференціального приводу перехід від s до s' визначається відповідними рівняннями руху. Однак результат не завжди є детермінованим, стохастичність у модель вносять такі фактори, як пробуксовування коліс, шум сенсорних даних та непередбачувана поведінка динамічних перешкод.



Функція винагороди $R(s, a)$ стимулює поведінку робота, яка сприяє досягненню бажаної мети, уникаючи небажаних наслідків.

Коефіцієнт дисконтування γ визначає, наскільки велика увага приділяється негайній та довгостроковій ефективності навігації. Наприклад, γ близьке до 1 гарантує, що робот враховує всю траєкторію до цілі.

Мета агента RL полягає в пошуку оптимальної політики $\pi(a|s)$, що визначає ймовірність виконання дії в заданому стані s для максимізації очікуваної кумулятивної дисконтованої винагороди:

$$J(\pi) = E_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right],$$

де s_t і a_t — стан та дія на кроці t відповідно.

Якість політики π оцінюється за допомогою функції цінності стану $V_{\pi}(s)$ або функції цінності стану-дії $Q_{\pi}(s, a)$, які визначаються наступним чином:

$$V_{\pi}(s) = E_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) | s_0 = s \right]$$

$$Q_{\pi}(s, a) = E_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) | s_0 = s, a_0 = a \right]$$

Оптимальна політика π_* максимізує функції $V_{\pi}(s)$ і $Q_{\pi}(s, a)$, та задовольняє рівняння оптимальності Беллмана:

$$Q^*(s, a) = R(s, a) + \gamma E_{s' \sim P} \left[\max_{a' \in A} Q^*(s', a') \right],$$

Рівняння оптимальності Беллмана є теоретичною основою для багатьох алгоритмів, спрямованих на пошук оптимальних політик у задачах послідовного прийняття рішень. Дане рівняння дозволяє розкласти задачу довгострокової оптимізації на менші підзадачі, використовуючи рекурсивний характер задачі.

Модель соціально-адаптивної навігації автономного мобільного робота. Розробка соціально-адаптивної навігації виходить за рамки традиційного планування шляху, в якому людина розглядається лише як динамічна перешкода. Мета нашого дослідження полягає в тому, щоб наділити робота здатністю рухатись у спосіб, який є не тільки безпечним, але й зрозумілим, передбачуваним та психологічно комфортним для людини. Для досягнення цієї мети необхідно інтегрувати в модель навігації декілька ключових принципів, що базуються на психології людської взаємодії.

Людський мозок постійно намагається прогнозувати рух об'єктів у навколишньому середовищі для забезпечення власної безпеки. Плавні, безперервні траєкторії мобільного робота підвищують передбачуваність його поведінки для людини. Навпаки, хаотичні рухи змушують людину постійно переоцінювати наміри робота, що збільшує її рівень стресу. Окрім динамічних характеристик руху, не менш важливим є його просторовий аспект, зокрема, траєкторія руху робота відносно людини. Одним із найбільш некомфортних сценаріїв для людини є зустрічний рух робота у фронтальній зоні людини. Робот, який рухається назустріч у фронтальній зоні людини, змушує її активно реагувати на його дії — зупинятися чи змінювати траєкторію. Така поведінка суперечить не лише комфортній взаємодії, але й фундаментальним нормам міжособистісної відстані, які формалізовані в теорії проксемики. Згідно з цією теорією, необґрунтоване вторгнення робота в інтимну та персональну зони людини викликає сильний дискомфорт і може сприйматися як загроза.



Таким чином, плавність руху робота, уникнення зустрічного руху у фронтальній зоні людини та дотримання соціальних дистанцій формують критерії соціально-адаптивної навігації. Їхній вибір ґрунтується на розумінні людської психології та соціальної динаміки, а комплексна інтеграція дозволяє роботу не просто уникати людей, а гармонійно співіснувати з ними у спільному просторі.

Для формалізації задачі навігації мобільного робота в динамічному середовищі з дотриманням норм соціальної взаємодії застосовуємо апарат марківських процесів прийняття рішень (MDP). Даний підхід дозволяє мобільному роботу навчатися оптимальній поведінці шляхом взаємодії з середовищем.

MDP визначено кортежем (S, A, P, R, γ) , де кожен елемент має наступне визначення в рамках даного дослідження.

Для створення математичної моделі потрібно визначити стани, дії та винагороди, які використовуватимуться автономним мобільним роботом для навчання.

Вектор стану S_t визначається як:

$$S_t = [L_t, d_g, \varphi_g, d_h, \varphi_h, v_{hx}, v_{hy}, \psi_h, v_r, \omega_r],$$

де

$L_t \in R^N$ — масив з N значень даних з лазерного сканера LIDAR;

d_g — відстань від робота до цілі;

φ_g — кут до цілі відносно напрямку робота;

d_h — відстань від робота до людини;

φ_h — кут до людини відносно напрямку робота;

v_{hx}, v_{hy} — складові вектора лінійної швидкості людини;

ψ_h — кут орієнтації людини;

v_r, ω_r — лінійна та кутова швидкість робота.

Простір дій (A) є неперервним. Дії передаються безпосередньо на контролер мобільного робота. Робот здійснює дії $a_t \in A \subset R^2$, які включають лінійну та кутову швидкості:

$$a_t = [v_t, w_t],$$

Функція винагороди $R(s_t, a_t)$ є ключовим елементом, який скеровує процес навчання. Вона розроблена як сума компонентів, які заохочують ефективне досягнення цілі та карають за небезпечну або соціально неприйнятну поведінку.

Функція винагороди формалізована як композиція декількох компонент:

$$R = R_{safety} + R_{efficiency} + R_{social}$$

1. R_{safety} — компонент відповідає за уникнення зіткнень з будь-якими перешкодами (статичними перешкодами або динамічними агентами). Даний компонент функції винагороди є пріоритетним, оскільки безпека є базовою вимогою до роботів, які взаємодіють з людьми. Зіткнення — це термінальний стан, який завершує епізод.

$$R_{safety} = R_{coll} + R_{obs_prox},$$

де R_{coll} — штраф за зіткнення зі статичною перешкодою або людиною, якщо робот наближається ближче, ніж на відстань зіткнення d_{coll} ;

R_{obs_prox} — штраф за наближення до перешкод ближче, ніж на відстань d_{warn} , якщо $d_{min} < d_{warn}$

$$R_{obs_prox} = -k_{obs_prox}(d_{warn} - d_{min}),$$

де d_{min} — мінімальна відстань лазерного сканера, k_{obs_prox} — коефіцієнт.

2. $R_{efficiency}$ — компонент функції винагороди, який відповідає за ефективність досягнення цілі.



$$R_{efficiency} = R_{goal} + R_{distance} + R_{step},$$

$R_{efficiency}$ містить такі складові: R_{goal} — позитивна винагорода за досягнення цілі, $R_{distance}$ — винагорода за наближення до цілі: позитивна, коли робот наближається до цільової точки, та негативна, якщо віддаляється.

$$R_{distance} = k_g \cdot (\Delta d_{g,t}) = k_g \cdot (d_{g,t-1} - d_{g,t}),$$

де $d_{g,t}$ — відстань до цільової точки в момент часу t , k_g — коефіцієнт.

R_{step} — негативна винагорода за кожен крок, яка стимулює робота до дії.

Компонент винагороди $R_{efficiency}$ забезпечує функціональність робота, але може конфліктувати з соціальним комфортом, якщо коротший шлях проходить через особистий простір людини.

3. R_{social} — компонент винагороди, який моделює соціально-адаптивну поведінку робота при взаємодії з людьми

$$R_{social} = R_{prox_int} + R_{prox_pers} + R_{angular} + R_{frontal}$$

Дана винагорода містить такі компоненти:

R_{prox_int} — штраф за вторгнення робота в інтимну проксемічну зону. Він нараховується за умови $d_h < d_{int}$

$$R_{prox_int} = -k_{int} \left(\frac{d_{int} - d_h}{d_{int} - d_{coll}} \right)$$

$R_{personal}$ — штраф за перебування в особистій проксемічній зоні, який нараховується за умови $d_{int} < d_h < d_{pers}$

$$R_{prox_pers} = -k_{pers} \left(\frac{d_{pers} - d_h}{d_{pers} - d_{int}} \right)$$

$R_{angular}$ — штраф за високу кутову швидкість робота в соціальній зоні, заохочує робота до плавних та передбачуваних рухів, нараховується, якщо $\omega_r > \omega_{max}$

$$R_{angular} = -k_a |\Delta \omega|$$

$R_{frontal}$ — штраф за фронтальне наближення робота до людини у соціальній зоні, нараховується за умови, якщо $d_{pers} < d_h < d_{social}$

$$R_{frontal} = -k_f \left(\frac{d_{social} - d_h}{d_{social} - d_{pers}} \right)$$

де d_{social} , d_{pers} , d_{int} — радіуси соціальної, особистої та інтимної зон відповідно, d_h — відстань між роботом і людиною, k_{int} , k_{pers} , k_f , k_l , k_a — коефіцієнти.

Компонент R_{social} функції винагороди безпосередньо конфліктує з ефективністю, оскільки соціально прийнятний шлях може бути довшим або повільнішим.

Коефіцієнт дисконтування $\gamma \in [0, 1)$ визначає стратегію поведінки мобільного робота. Цей гіперпараметр визначає часовий горизонт планування та встановлює баланс між цінністю негайних винагород та винагород, які будуть отримані у віддаленому майбутньому. Значення γ , яке наближається до 1, змушує агента враховувати довгострокові наслідки своїх дій, що є критично важливим для задач навігації, де найкоротший шлях до миттєвої винагороди може призвести до небажаних станів (наприклад, потрапляння в локальний мінімум або небезпечне зближення з людиною).

Агент навчається оптимальній політиці на основі досвіду, отриманого шляхом взаємодії з середовищем.



Для практичної реалізації запропонованої моделі застосовуються алгоритми навчання з підкріпленням, які здатні ефективно працювати в складних, неперервних просторах станів і дій. Зокрема, Proximal Policy Optimization (PPO) вирізняється своєю стабільністю та ефективністю використання даних. Водночас алгоритм Soft Actor-Critic (SAC) демонструє високу продуктивність у неперервних просторах дій та сприяє кращому дослідженню середовища за рахунок максимізації ентропії, що може призвести до знаходження більш робастних стратегій поведінки.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У процесі проведеного дослідження було розроблено модель соціально-адаптивної навігації мобільного робота на основі навчання з підкріпленням.

Подальші дослідження можуть бути спрямовані на розробку та впровадження методів навчання соціально-адаптивній поведінці робота, які дадуть змогу автономному мобільному роботу краще пристосовуватися до динаміки руху людей. Важливим напрямком подальших досліджень є вдосконалення моделі, зокрема, підвищення точності прогнозування та реакції автономного мобільного робота на непередбачувані зміни руху людей.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Kruse, T., Pandey, A. K., Alami, R., & Kirsch, A. (2013). Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12), 1726–1743. <https://doi.org/10.1016/j.robot.2013.05.007>
2. Yin, X., Yulian, C., Cheng, M., Liu, W., Dong, W., & Yao, D. (2024). Path planning of mobile robot based on improved D* Lite_TEB algorithm. *IET Conference Proceedings*, 2023(49), 26–31. <https://doi.org/10.1049/icp.2024.3623>
3. Li, Y., Jin, R., Xu, X., Qian, Y., Wang, H., Xu, S., & Wang, Z. (2022). A mobile robot path planning algorithm based on improved A* algorithm and dynamic window approach. *IEEE Access*, 1. <https://doi.org/10.1109/access.2022.3179397>
4. Forer, S., Banisetty, S. B., Yliniemi, L., Nicolescu, M., & Feil-Seifer, D. (2018). Socially-aware navigation using non-linear multi-objective optimization. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 1–8). IEEE. <https://doi.org/10.1109/iros.2018.8593825>
5. Teso-Fz-Betoño, D., Zulueta, E., Fernandez-Gamiz, U., Saenz-Aguirre, A., & Martinez, R. (2019). Predictive dynamic window approach development with artificial neural fuzzy inference improvement. *Electronics*, 8(9), 935. <https://doi.org/10.3390/electronics8090935>
6. Kang, S., Yang, S., Kwak, D., Jargalbaatar, Y., & Kim, D. (2024). Social type-aware navigation framework for mobile robots in human-shared environments. *Sensors*, 24(15), 4862. <https://doi.org/10.3390/s24154862>
7. Kivrak, H., Cakmak, F., Kose, H., & Yavuz, S. (2020). Social navigation framework for assistive robots in human inhabited unknown environments. *Engineering Science and Technology, an International Journal*. <https://doi.org/10.1016/j.jestch.2020.08.008>
8. Faust, A., Oslund, K., Ramirez, O., Francis, A., Tapia, L., Fiser, M., & Davidson, J. (2018). PRM-RL: Long-range robotic navigation tasks by combining reinforcement learning and sampling-based planning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1–9). IEEE. <https://doi.org/10.1109/icra.2018.8461096>
9. Chen, Y. F., Liu, M., Everett, M., & How, J. P. (2017). Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning. In *2017 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1–8). IEEE. <https://doi.org/10.1109/icra.2017.7989037>
10. Kahn, G., Villafior, A., Ding, B., Abbeel, P., & Levine, S. (2018). Self-supervised deep reinforcement learning with generalized computation graphs for robot navigation. In *2018 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1–8). IEEE. <https://doi.org/10.1109/icra.2018.8460655>



11. Chen, Y. F., Everett, M., Liu, M., & How, J. P. (2017). Socially aware motion planning with deep reinforcement learning. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 1–8). IEEE. <https://doi.org/10.1109/iros.2017.8202312>
12. Hanenko, L. D., & Zhebka, V. V. (2023). Analytical review of navigation issues of mobile robots in indoor environment. *Telecommunication and Information Technologies*, 80(3). <https://doi.org/10.31673/2412-4338.2023.038087>
13. Ngo, H. Q. T., Le, V. N., Thien, V. D. N., Nguyen, T. P., & Nguyen, H. (2020). Develop the socially human-aware navigation system using dynamic window approach and optimize cost function for autonomous medical robot. *Advances in Mechanical Engineering*, 12(12), 168781402097943. <https://doi.org/10.1177/1687814020979430>
14. Wang, Y., Yu, J., Kong, Y., Sun, L., Liu, C., Wang, J., & Chi, W. (2024). Socially adaptive path planning based on generative adversarial network. *IEEE Transactions on Intelligent Vehicles*, 1–13. <https://doi.org/10.1109/tiv.2024.3478219>
15. Singamaneni, P. T., Bachiller-Burgos, P., Manso, L. J., Garrell, A., Sanfeliu, A., Spalanzani, A., & Alami, R. (2024). A survey on socially aware robot navigation: Taxonomy and future challenges. *The International Journal of Robotics Research*. <https://doi.org/10.1177/02783649241230562>
16. Li, K., Lu, Y., & Meng, M. Q. H. (2021). Human-aware robot navigation via reinforcement learning with hindsight experience replay and curriculum learning. In *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)* (pp. 1–7). IEEE. <https://doi.org/10.1109/robio54168.2021.9739519>
17. Ngo, H. Q. T., Le, V. N., Thien, V. D. N., Nguyen, T. P., & Nguyen, H. (2020). Develop the socially human-aware navigation system using dynamic window approach and optimize cost function for autonomous medical robot. *Advances in Mechanical Engineering*, 12, 168781402097943. <https://doi.org/10.1177/1687814020979430>
18. Daza, M., Barrios-Aranibar, D., Diaz-Amado, J., Cardinale, Y., & Vilasboas, J. (2021). An approach of social navigation based on proxemics for crowded environments of humans and robots. *Micromachines*, 12(2), 193. <https://doi.org/10.3390/mi12020193>
19. Kang, S., Yang, S., Kwak, D., Jargalbaatar, Y., & Kim, D. (2024). Social type-aware navigation framework for mobile robots in human-shared environments. *Sensors*, 24(15), 4862. <https://doi.org/10.3390/s24154862>
20. Hanenko, L. D., & Zhebka, V. V. (2024). Application of reinforcement learning methods for path planning of mobile robots. *Telecommunication and Information Technologies*, (1), 16–25. <https://doi.org/10.31673/2412-4338.2024.011625>

**Liudmyla Hanenko**

PhD student at the Department of Computer Engineering

State University of Information and Communication Technologies, Kyiv, Ukraine

ORCID ID: 0000-0003-2219-8196

hanenkoliudmyla@gmail.com**Viktoriia Zhebka**

Doctor of Technical Sciences, Associate Professor

State University of Information and Communication Technologies, Kyiv, Ukraine

ORCID ID: 0000-0003-4051-1190

viktoria_zhebka@ukr.net.com

MODEL OF SOCIAL-ADAPTIVE NAVIGATION OF A MOBILE ROBOT USING REINFORCEMENT LEARNING METHODS

Abstract. Classic trajectory planning algorithms, despite their effectiveness in static environments, demonstrate significant limitations when integrated into a dynamic social environment. The main drawback is the inability to interpret human movement in real time, which leads to unpredictable and potentially dangerous maneuvers. In response to these limitations, reinforcement learning (RL) methods have gained considerable popularity. This paradigm allows an autonomous mobile robot to independently form an optimal behavior strategy through direct interaction with the environment and receiving feedback in the form of rewards or penalties. This study focuses on reinforcement learning methods and social behavior models with the aim of developing safe, effective, and socially adaptive navigation for autonomous mobile robots. This paper presents a model in which the key feature is a comprehensive approach to shaping agent behavior. The proposed model takes into account not only basic tasks, goal achievement, and avoidance of physical obstacles, but also important aspects of social interaction. The scientific novelty of the work lies in the development of a multicomponent reward function that integrates rewards for reaching the target point, avoiding collisions with dynamic and static objects, and purposefully encourages the agent to adhere to socially acceptable norms. In this way, the robot learns not only to avoid people, but to do so in a way that is intuitive and comfortable for them. The ultimate goal of the research is to create a navigation agent that is not only safe, but also socially intelligent. This is a step towards the full integration of autonomous robotic systems into everyday human environments, as successful coexistence requires not only physical safety, but also psychological comfort and intuitive understanding of the robot's behavior.

Keywords: information technology; modeling; machine learning methods; reinforcement learning methods; autonomous mobile robots mobile robot navigation.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Kruse, T., Pandey, A. K., Alami, R., & Kirsch, A. (2013). Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12), 1726–1743. <https://doi.org/10.1016/j.robot.2013.05.007>
2. Yin, X., Yulian, C., Cheng, M., Liu, W., Dong, W., & Yao, D. (2024). Path planning of mobile robot based on improved D* Lite_TEB algorithm. *IET Conference Proceedings*, 2023(49), 26–31. <https://doi.org/10.1049/icp.2024.3623>
3. Li, Y., Jin, R., Xu, X., Qian, Y., Wang, H., Xu, S., & Wang, Z. (2022). A mobile robot path planning algorithm based on improved A* algorithm and dynamic window approach. *IEEE Access*, 1. <https://doi.org/10.1109/access.2022.3179397>
4. Forer, S., Banisetty, S. B., Yliniemi, L., Nicolescu, M., & Feil-Seifer, D. (2018). Socially-aware navigation using non-linear multi-objective optimization. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 1–8). IEEE. <https://doi.org/10.1109/iros.2018.8593825>
5. Teso-Fz-Betoño, D., Zulueta, E., Fernandez-Gamiz, U., Saenz-Aguirre, A., & Martinez, R. (2019). Predictive dynamic window approach development with artificial neural fuzzy inference improvement. *Electronics*, 8(9), 935. <https://doi.org/10.3390/electronics8090935>



6. Kang, S., Yang, S., Kwak, D., Jargalbaatar, Y., & Kim, D. (2024). Social type-aware navigation framework for mobile robots in human-shared environments. *Sensors*, 24(15), 4862. <https://doi.org/10.3390/s24154862>
7. Kivrak, H., Cakmak, F., Kose, H., & Yavuz, S. (2020). Social navigation framework for assistive robots in human inhabited unknown environments. *Engineering Science and Technology, an International Journal*. <https://doi.org/10.1016/j.jestch.2020.08.008>
8. Faust, A., Oslund, K., Ramirez, O., Francis, A., Tapia, L., Fiser, M., & Davidson, J. (2018). PRM-RL: Long-range robotic navigation tasks by combining reinforcement learning and sampling-based planning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1–9). IEEE. <https://doi.org/10.1109/icra.2018.8461096>
9. Chen, Y. F., Liu, M., Everett, M., & How, J. P. (2017). Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning. In *2017 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1–8). IEEE. <https://doi.org/10.1109/icra.2017.7989037>
10. Kahn, G., Villafior, A., Ding, B., Abbeel, P., & Levine, S. (2018). Self-supervised deep reinforcement learning with generalized computation graphs for robot navigation. In *2018 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1–8). IEEE. <https://doi.org/10.1109/icra.2018.8460655>
11. Chen, Y. F., Everett, M., Liu, M., & How, J. P. (2017). Socially aware motion planning with deep reinforcement learning. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 1–8). IEEE. <https://doi.org/10.1109/iros.2017.8202312>
12. Hanenko, L. D., & Zhebka, V. V. (2023). Analytical review of navigation issues of mobile robots in indoor environment. *Telecommunication and Information Technologies*, 80(3). <https://doi.org/10.31673/2412-4338.2023.038087>
13. Ngo, H. Q. T., Le, V. N., Thien, V. D. N., Nguyen, T. P., & Nguyen, H. (2020). Develop the socially human-aware navigation system using dynamic window approach and optimize cost function for autonomous medical robot. *Advances in Mechanical Engineering*, 12(12), 168781402097943. <https://doi.org/10.1177/1687814020979430>
14. Wang, Y., Yu, J., Kong, Y., Sun, L., Liu, C., Wang, J., & Chi, W. (2024). Socially adaptive path planning based on generative adversarial network. *IEEE Transactions on Intelligent Vehicles*, 1–13. <https://doi.org/10.1109/tiv.2024.3478219>
15. Singamaneni, P. T., Bachiller-Burgos, P., Manso, L. J., Garrell, A., Sanfeliu, A., Spalanzani, A., & Alami, R. (2024). A survey on socially aware robot navigation: Taxonomy and future challenges. *The International Journal of Robotics Research*. <https://doi.org/10.1177/02783649241230562>
16. Li, K., Lu, Y., & Meng, M. Q. H. (2021). Human-aware robot navigation via reinforcement learning with hindsight experience replay and curriculum learning. In *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)* (pp. 1–7). IEEE. <https://doi.org/10.1109/robio54168.2021.9739519>
17. Ngo, H. Q. T., Le, V. N., Thien, V. D. N., Nguyen, T. P., & Nguyen, H. (2020). Develop the socially human-aware navigation system using dynamic window approach and optimize cost function for autonomous medical robot. *Advances in Mechanical Engineering*, 12, 168781402097943. <https://doi.org/10.1177/1687814020979430>
18. Daza, M., Barrios-Aranibar, D., Diaz-Amado, J., Cardinale, Y., & Vilasboas, J. (2021). An approach of social navigation based on proxemics for crowded environments of humans and robots. *Micromachines*, 12(2), 193. <https://doi.org/10.3390/mi12020193>
19. Kang, S., Yang, S., Kwak, D., Jargalbaatar, Y., & Kim, D. (2024). Social type-aware navigation framework for mobile robots in human-shared environments. *Sensors*, 24(15), 4862. <https://doi.org/10.3390/s24154862>
20. Hanenko, L. D., & Zhebka, V. V. (2024). Application of reinforcement learning methods for path planning of mobile robots. *Telecommunication and Information Technologies*, (1), 16–25. <https://doi.org/10.31673/2412-4338.2024.011625>

