

Сарнацький В.В.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Баклан І.В.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ВПЛИВ СОЦІОДЕМОГРАФІЧНОЇ ГЕТЕРОГЕННОСТІ НА ОПТИМАЛЬНУ СТРАТЕГІЮ ВПРОВАДЖЕННЯ КАРАНТИННИХ ЗАХОДІВ

У роботі описані результати дослідження оптимальних стратегій впровадження карантинних заходів до модельованих середовищ із різною соціодемографічною параметризацією.

У межах цієї роботи було реалізоване середовище моделювання розповсюдження епідемії у певній країні. Ця модель може бути гнучко параметризована заданням таких параметрів, як: кількість мешканців країни, їх віковий розподіл, розподіл розміру родин, кількість закладів навчання та роботи, розподіл приналежності кожного мешканця до закладу навчання та роботи в залежності від віку. За допомогою такої параметризації були побудовані моделі деяких європейських країн. Розповсюдження епідемії моделюється методом реєстрації контактів. Цей метод полягає у симуляції переміщень кожного мешканця країни протягом часу та його контактів з іншими мешканцями. Такий підхід до моделювання розповсюдження епідемії, хоча і є більш вимогливим до обчислювальних ресурсів, здатний відображати складні стохастичні процеси, приманні реальному світу.

На основі розробленої моделі було побудоване середовище навчання інтелектуального агента, сформульоване як Марковський процес прийняття рішень. В якості доступних агенту дій були обрані три дії: повний карантин у всіх закладах, за якого всі мешканці контактують лише зі своїми родичами; скасування карантину та бездіяльність. Винагорода за обрану дію була сформована з двох складників: винагорода за економічний стан та винагорода за епідеміологічний стан. В якості стану агент спостерігає за процентом кількості інфікованих мешканців за останні декілька днів.

Модель прийняття рішень агента побудована на основі повнозв'язної штучної нейронної мережі, що приймає на вхід поточний стан середовища. Вагові коефіцієнти нейронної мережі підбираються за допомогою алгоритму глибокого Q-навчання. Для валідації окремо навченого агента його ефективність була порівняна з ефективностями трьох інших стратегій: стратегії повного карантину, стратегії ігнорування і стратегії випадкових дій. Також агенти, що були навчені на середовищах, параметризованих розподілами однієї країни, були протестовані на середовищах інших країн. Результати порівняння показали, що у більшості випадків агент, що був навчений на певній країні, показує більшу ефективність, ніж інші агенти. З цього можна зробити висновок, що соціодемографічні розподіли відіграють важливу роль у процесі пошуку оптимальної стратегії.

Ключові слова: навчання з підкріпленням, інтелектуальні агенти, штучний інтелект, епідеміологія.

Постановка проблеми. Вплив пандемії COVID-19, спричиненої коронавірусом SARS-CoV-2 [1, с. 1], показав, що відсутність оптимальної стратегії впровадження карантинних заходів здатна вплинути як на економічний, так і на епідеміологічний стан ураженої країни [2, с. 7], а також на психологічний стан її населення [3, с. 1; 4, с. 3]. Результати моделювання та аналізу показали, що впровадження своєчасних карантинних засобів сприяє «згладжуванню» кривої захворювань, що дає змогу вчасно надати медичну допомогу хворим та оминати ситуацію з нестачею медичних

ресурсів. Проте визначення критеріїв «своєчасності» є нетривіальною задачею і потребує експертної оцінки епідеміологів, які, будучи людьми, здатні на помилки, що можуть стати причиною зайвих людських жертв та удару по економіці країни. Тому було вирішено дослідити специфіку використання інтелектуальних агентів, основним завданням яких буде контроль за карантинними заходами у середовищі моделювання.

Аналіз останніх досліджень і публікацій. Через те, що в основі роботи лежить середовище моделювання, насамперед була проаналізована

література щодо моделювання розповсюдження епідемій. Серед опрацьованих робіт значну частину займають праці, основані на компартментних моделях. Цей тип моделей ділить модельоване суспільство на декілька епідеміологічних груп та моделює зміни у складі цих груп протягом часу. Першою такою моделлю є модель SIR [5, с. 1], що передбачає наявність трьох таких груп: S (Susceptible) – група людей, що не мають імунітету до хвороби та можуть бути інфікованими; I (Infected або Infectious) – група людей, що були інфіковані та здатні інфікувати інших; R (Recovered або Removed) – група людей, що перехворіла та виробила імунітет до хвороби (у деяких роботах до цієї групи відносять загинувших через хворобу). Окрім SIR існує також модель SIS, що передбачає перехід членів групи I до групи S та їх повторне інфікування. Модель SEIR [6, с. 3] передбачає додаткову групу E (Exposed), члени якої були у контакті з інфікованими, проте не мають симптомів. Представники цієї групи можуть інфікувати групу S, проте з нижчим ступенем ймовірності (інколи рівним 0) у порівнянні з представниками групи I. Модель SEIRS [6, с. 3] є модифікацією моделі SEIR, що передбачає втрату імунітету через певний час та можливість повторного інфікування.

Для моделювання переходів представників однієї групи до іншої зміни у кількості представників кожної групи з часом моделюються системою диференціальних рівнянь. Аналіз таких систем із калібруванням моделей на реальних соціодемографічних даних дав змогу дослідити вплив різних видів карантинних заходів на динаміку розповсюдження захворювання [7, с. 1; 8, с. 14; 9, с. 15].

В останні роки набирають популярність так звані агент-орієнтовані підходи, що моделюють кожного представника групи окремо [10, с. 1]. Такий тип моделей дозволяє урахувати складні стохастичні процеси, що притаманні реальному світу, і здатен більш точно моделювати процес розповсюдження епідемій [11, с. 201; 12, с. 5].

Аналіз останніх досліджень і публікацій показав, що ідея використати інтелектуальні агенти для розв'язання поставленої задачі не є новою. Так, у [13, с. 2] були показані дослідження роботи агента, що контролював упровадження карантинних заходів у модельованому середовищі. Хоча модель розповсюдження епідемії були занадто спрощена, автори доповідають, що розроблений метод можна застосовувати з будь-якими моделями. У [14, с. 7] представлені

результати аналізу роботи складнішого агента, в основі якого лежить рекурентна нейронна мережа, що аналізує статистику захворювань і робить вибір з трьох альтернатив: скасувати усі карантинні впровадження, ввести 25% та 75% карантин. Результати показали, що агент здатний контролювати розповсюдження епідемії, тримаючи баланс між показниками епідеміологічного та економічного стану країни.

Постановка завдання. У результаті проведеного аналізу літератури не було знайдено досліджень щодо використання інтелектуальних агентів, навчених методами навчання з підкріпленням на агент-орієнтованих середовищах моделювання з соціодемографічною параметризацією. Дослідження у цій сфері є важливим, оскільки соціодемографічна параметризація може мати значний вплив на оптимальні стратегії впровадження карантинних заходів, а сучасні обчислювальні ресурси дозволяють відмовитися від класичних моделей розповсюдження епідемії на користь агент-орієнтованих.

Виклад основного матеріалу дослідження.

Модель розповсюдження епідемії.

У роботі в якості середовища для навчання та тестування агентів була вибрана індивідуальна модель, що є спрощеною версією моделі, описаної в [10]. Центральним об'єктом моделювання є країна з населенням P та трьома видами місць зустрічі жителів: житлові будинки, місця навчання та роботи, множини яких представлені як H, S, W , а їх кількість задана як: $|H| = \alpha_h \cdot |P|$, $|S| = \alpha_s \cdot |P|$, $|W| = \alpha_w \cdot |P|$ відповідно. Кожен мешканець міста $p \in P$ має вік $ap \sim Page$ та родину розміром $fs, p \sim Hsize$ з якою він проживає у будинку $hp \in H$. Також, в залежності від віку, мешканці можуть мати місце роботи та навчання відповідно до розподілів $w_p \sim P_w(ap)$, $sp \sim P_s(ap)$, $w \in W \cup \{\emptyset\}$, $s \in S \cup \{\emptyset\}$. Розподіли P_w, P_s задані так, що для кожного місця роботи на навчання кількість відповідних йому мешканців відповідає розподілам $Wsize$ та $Ssize$. Кожна установа $m \in W \cup S$ у час t може бути зачинена на карантин. У такому разі, величина $C_m(t) = 0$, інакше – $C_m(t) = 1$.

Кожен мешканець має стан інфікування $I_p(t) \in \{0, 1, 2, 3\}$ у час t , що відповідають моделі SEIRS (Рис. 1) наступним чином: 0 – Susceptible, 1 – Exposed, 2 – Infectious, 3 – Recovered. Перехід від стану 0 до стану 1 відбувається у разі інфікування, а переходи $1 \rightarrow 2$, $2 \rightarrow 3$, $3 \rightarrow 0$ відбуваються після проходження часу te, i , $te, i + ti, r$, $te, i + ti, r + tr, s$ рахуючи від часу інфікування te, p відповідно.

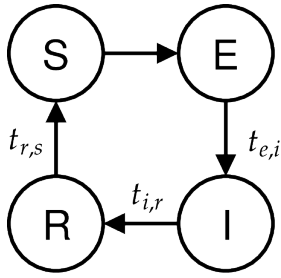


Рис. 1. Модель SEIRS та переходи між станами

Симуляція виконується з періодом в один день. Протягом цього дня кожен мешканець перебуває у наступних установах протягом наведеного інтервалу часу:

- $hp, [0, t_{wakeup})$;
- Якщо $(wp \in W \wedge C_{wp}(t) = 0)$, $(sp \notin S \vee C_{sp}(t) = 1)$: $wp, [t_{ws}, t_{we})$; $hp, [t_{we} + 1, 24)$;
- Якщо $(wp \notin W \vee C_{wp}(t) = 1)$, $(sp \notin S \wedge C_{sp}(t) = 0)$: $sp, [t_{ss}, t_{se})$; $hp, [t_{se} + 1, 24)$;
- Якщо $(wp \in W \wedge C_{wp}(t) = 0)$, $(sp \notin S \wedge C_{sp}(t) = 0)$: $wp, [t_{ws}, t_{we})$; $sp, [t_{ess}, t_{ese})$; $hp, [t_{ese} + 1, 24)$;
- Якщо $(wp \notin W \vee C_{wp}(t) = 1)$, $(sp \notin S \vee C_{sp}(t) = 1)$: $hp, [t_{wakeup}, 24)$;

Під кінець дня кожний мешканець p для якого $I_p(t) = 0$ інфікується ($1 \rightarrow I_p(t + 1)$) з ймовірністю:

$$1 - \prod_{p' \in P, I_{p'} \in \{1, 2\}, t=0, M(p, t) = M(p', t)}^{24} (1 - \alpha_{I_p}^{C_{M(p, t)} T(M(p, t))}), \quad (1)$$

де $M(p, t) : P \times \mathbb{N} \rightarrow H \cup W \cup S \cup \{\emptyset\}$ – функція, що для заданого мешканця p і часу t повертає місце, де він перебував у цей час. Через те, що модель може бути параметризована так, що буде існувати момент часу коли мешканець не буде перебувати ні в якій установі – функція у такий випадок повертає \emptyset . $T(m) : H \cup W \cup S \cup \{\emptyset\} \rightarrow \mathbb{R}$ – функція, що відображає установу в ступінь інфікування перебуваючи в ній. Параметр α_{I_p} відображає коефіцієнт ймовірності інфікування носієм зі станом I_p' .

Подальші результати отримані при наступних значеннях параметрів моделі:

- $t_{wakeup} = 7$;
- $t_{ws} = 9$;
- $t_{we} = 17$;
- $t_{ss} = 9$;
- $t_{se} = 15$;
- $t_{ess} = 18$;
- $t_{ese} = 20$;
- $T(h) = 2 \cdot 10^{-4}, \forall h \in H$;
- $T(s) = 2 \cdot 10^{-4}, \forall s \in S$;
- $T(w) = 1 \cdot 10^{-4}, \forall w \in W$;
- $t_{e,i} = 14$;
- $t_{i,r} = 14$;
- $t_{r,s} = 10^{12}$.

Соціодемографічні дані.

Для отримання значень параметрів $\alpha_h, \alpha_s, \alpha_w, Page, P_w, P_s, Hsize, Wsize, Ssize$ була проаналі-

зована відкрита база статистичних даних надана Statistical Office of the European Communities (Eurostat), зокрема:

- Families by type, size and NUTS 3 region – для побудови розподілу $Hsize$;
- Population on 1 January by age and sex – для побудови розподілу $Page$;
- Pupils and students enrolled by education level, sex and age – для побудови розподілів $P_s, Ssize$;
- Employment by sex, age and NUTS 2 regions (1 000) – для побудови розподілів $P_w, Wsize$.

У разі відсутності певних даних для деякої країни – вони були замінені на усереднені дані по Європейському регіону (EU28).

Середа моделювання як Марковський процес прийняття рішень.

Для спрощення подальшої нотації введемо наступну функцію:

$$\begin{cases} \delta(P) = 1, P \\ \delta(P) = 0, -P \end{cases} \quad (2)$$

Через те, що для використання методу навчання з підкріпленням середовище має бути представлене як Марковський процес прийняття рішень [15, с. 47; 20, с. 1], були введені такі визначення:

- множина дій $A = \{0, 1, 2\}$, де 0 відповідає відсутності дії (залишити всі поточні карантинні впровадження та не вводити нових), 1 – повний карантин ($\forall m, 0 \rightarrow C_m$), 2 – скасування карантину ($\forall m, 1 \rightarrow C_m$);
- стан S_t у момент часу t заданий формулою 3;
- винагорода R_{t+1} за дію A_t задана формулою 4.

$$\forall t, S_t = p \in P \{h_p, w_p, s_p, t_{e,p}\} \cup \{C_m(t) | m \in W \cup S\} \quad (3)$$

$$R_{t+1} = R_{t+1}^{(e)} + \lambda R_{t+1}^{(i)}$$

$$R_t^{(i)} = \frac{-1}{p \vee \sum_{p \in P} \delta(I_p(t) = 1)}$$

$$R_t^{(e)} = \frac{1}{P \vee \sum_{p \in P} [S_H(p, t) + S_W(p, t) + S_S(p, t)]}$$

$$S_H(p, t) = \gamma_H \cdot \delta(I_p(t) = 1)$$

$$S_W(p, t) = \delta(w_p \neq \emptyset) \cdot (\gamma_{w'} + C(W_p, t)(\gamma_w - \gamma_{w'}))$$

$$S_S(p, t) = \delta(s_p \neq \emptyset) \cdot (\gamma_{s'} + C(S_p, t)(\gamma_s - \gamma_{s'})) \quad (4)$$

де $R_{t+1}^{(e)}$ – економічна винагорода, що показує стан економіки модельованої країни на момент часу $t + 1$; $R_{t+1}^{(i)}$ – епідеміологічна винагорода, що показує загальний епідеміологічний стан модельованої країни на момент часу $t + 1$; λ – ваговий коефіцієнт. Таке розбиття винагороди є цілком природним, бо використання в якості R_{t+1} лише одного з доданків приводить до тривіальної оптимальної стратегії: якщо $R_{t+1} = R_{t+1}^{(e)}$, оптимальна стратегія $\pi(s) = 2$ (повна бездіяльність), якщо $R_{t+1} = R_{t+1}^{(i)}$ оптимальна стратегія $\pi(s) = 1$ (повний карантин). Параметр λ має бути піді-

браний залежно від глобальної політики щодо компромісу між карантинними обмеженнями та їх впливу на економіку. Тому пошук оптимального його значення, так само, як і формулювання поняття «оптимальності» цього параметра, лежать за межами цієї роботи. Наведені у цій роботі результати відповідають $\lambda = 100$.

Опис агента.

Модель прийняття рішень, що лежить в основі агента, являє собою повнозв'язну штучну нейронну мережу, що приймає на вхід стан S_t та видає на виході q-значення кожної дії $q(s,a)$, що дорівнюють максимальній сумі можливих майбутніх винагород після обрання цієї дії. Таке формулювання, описане в [15], дозволяє агенту діяти з метою максимізації майбутньої винагороди, що є основою навчання з підкріпленням. Стан S_t і процес обчислення $q(s,a)$ задані таким чином:

$$\begin{aligned} \forall t, \bar{S}_t &\in R^D \\ \forall n = 1 \dots D, \bar{S}_{t_n} &= 0 \\ \forall n = 2 \dots D, \bar{S}_{t_n} &= \bar{S}_{t_{n-1}} \\ \bar{S}_{t_1} &= \frac{1}{|P|} \sum_{p \in P} \delta(I_p(t) = 2) \\ a_0 &= \bar{S}_t \\ x_i &= a_{i-1} \times W_{i-1} + b_{i-1} \\ a_i &= ReLU(BN(x_i)) \\ q(s,a) &= x_{L,a} \end{aligned} \tag{5}$$

де W_i, b_i – вагові коефіцієнти i -го шару нейронної мережі; $ReLU$ – функція активації ReLU [16, с. 1]; BN – функція batch-нормалізації [17, с. 1].

У процесі навчання агента його вагові коефіцієнти підбираються за допомогою алгоритму Deep

Q-learning with experience replay [18, с. 1] з використанням як методу оптимізації алгоритму Adam [19, с. 1].

Тестування.

Тестування було проведено на середовищах, побудованих на основі соціодемографічних даних таких країн, як: Болгарія, Данія, Латвія, Нідерланди, Польща, Швеція, Словенія, Велика Британія. Агент був побудований з $L = 1$.

На рис. 2 представлені результати валідації навчених агентів на описаному середовищі. Кожна комірка таблиці містить середню нормалізовану винагороду, що обчислена за формулою 6.

$$E_{c^1, c^2} = \frac{\sum_t R_t^{(c^2)}}{\sum_t R_t^{(c^1)}}, \tag{6}$$

де c^1 – країна тренування агента, c^2 – країна тестування агента, $R_t^{(c)}$ – винагорода у час t для середовища країни c .

Як можна побачити, 62.5% комірок містять величину, меншу за 1, що свідчить про те, що більшість результатів відповідають ситуації, коли агент, що навчався на середовищі певної країни, показує більшу кумулятивну винагороду, ніж агенти, що були натреновані на середовищах інших країн.

Через те, що єдина різниця між середовищами різних країн – це різні соціодемографічна статистика, можна зробити висновок, що під час пошуку оптимальної стратегії у впровадженні карантинних заходів слід враховувати соціодемографічну статистику цільової країни.

Було проведено додаткове порівняння натренованих агентів з агентами з такими стратегіями: $\pi(s) = 1, \pi(s) = 2, \pi(s) \sim U\{0, 2\}$. Результати наведені на рис. 3.

Висновки. У роботі викладені результати дослідження щодо використання інтелектуальних агентів для контролю за впровадженням карантинних заходів у середовищі моделювання розповсюдження захворювань з соціодемографічною параметризацією. Створене середовище моделювання було сформульоване як Марківський процес прийняття рішень, були введені поняття економічної та епідеміологічної винагороди. Були навчені інтелектуальні агенти на середовищах із параметризацією, аналогічною країнам Європейського союзу. З результатів валідації агентів можна зробити висновок, що навчені представленим методом агенти показують більшу ефективність, ніж тривіальні стратегії повного карантину, ігнорування та випадкового прийняття рішень, а також що соціодемографічна параметризація має вплив на ефективність роботи агента.

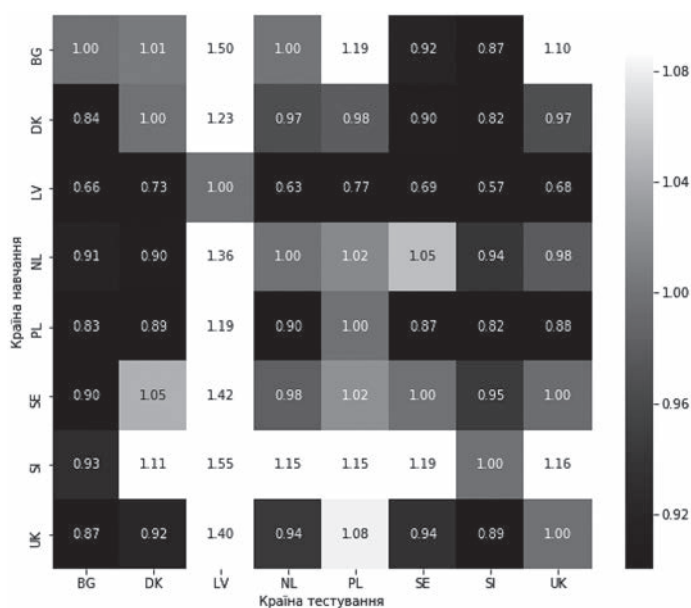


Рис. 2. Результати тестування

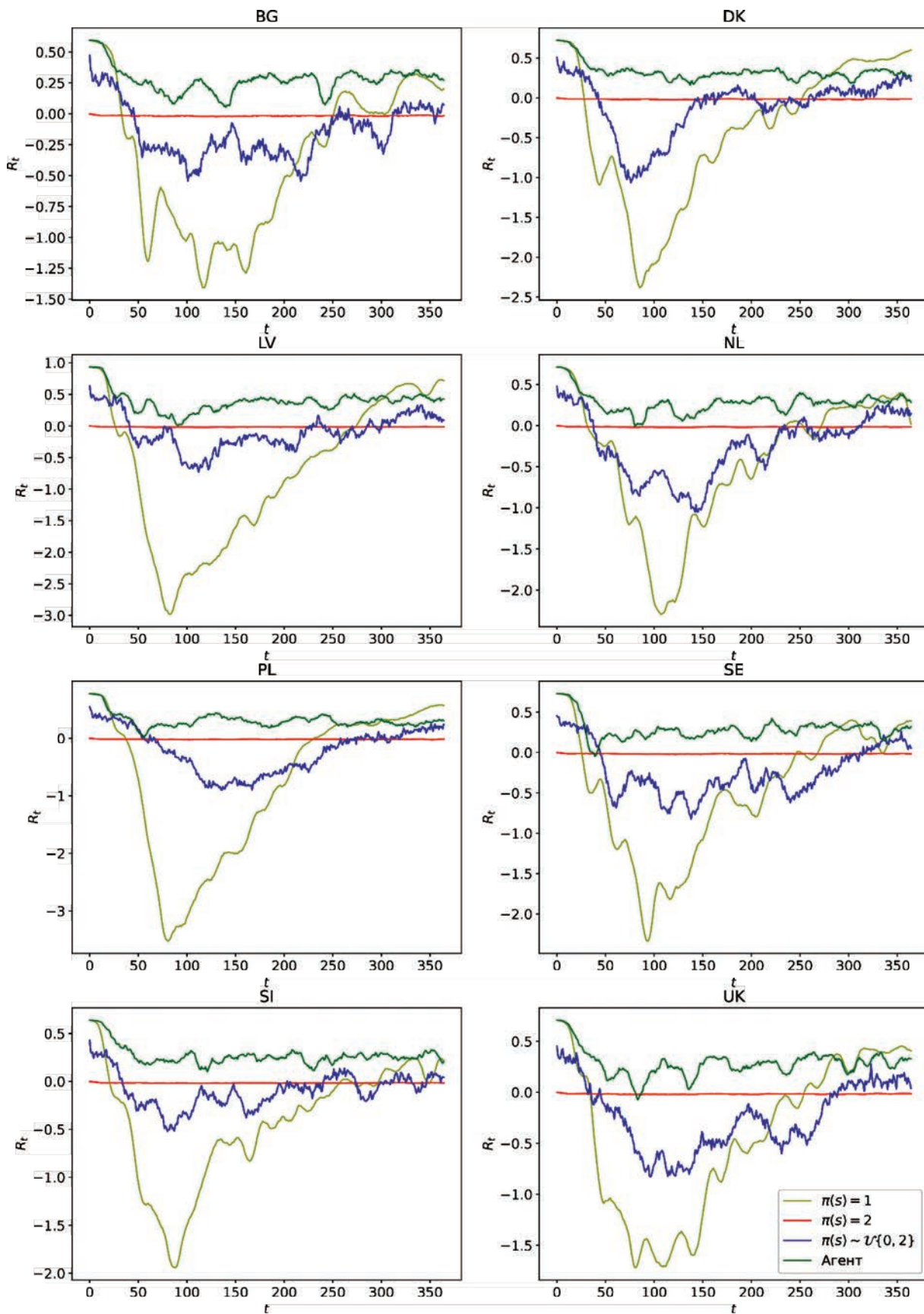


Рис. 3. Винагорода R_t протягом часу.
Зверху кожного графіка зазначена країна навчання та тестування агента

Список літератури:

1. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature microbiology*. 2020. T. 5. № 4. 536 с.
2. Dandekar, R. Quantifying the effect of quarantine control in COVID-19 infectious spread using machine learning. *medRxiv*. 2020.
3. Altuntas, F. The effect of COVID-19 pandemic on domestic tourism: A de-matel method analysis on quarantine decisions. *International Journal of Hospitality Management*. 2021. T. 92. 102719 с.
4. Yunhe W. The impact of quarantine on mental health status among general population in China during the COVID-19 pandemic. *Molecular psychiatry*. 2021. С. 1–10.
5. Harko, T. Exact analytical solutions of the susceptible-infected-recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation*. 2014. T. 236. С. 184–194.
6. Hethcote, H.W. Some epidemiological models with nonlinear incidence. *Journal of Mathematical Biology*. 1991. T. 29. № 3. С. 271–287.
7. Radulescu, A. Management strategies in a SEIR-type model of COVID-19 community spread. *Scientific reports*. 2020. T. 10. № 1. С. 1–16.
8. Fernando E Alvarez. A simple planning problem for COVID-19 lockdown. *National Bureau of Economic Research*. 2020.
9. Ng, W.L. To lockdown? When to peak? Will there be an end? a macroeconomic analysis on COVID-19 epidemic in the United States. *Journal of Macroeconomics*. 2020. T. 65. 103230 с.
10. Merler, S. The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proceedings of the Royal Society B: Biological Sciences*. 2010. T. 277. № 1681. С. 557–565.
11. Railsback, S.F. Agent-based and individual-based modeling: a practical introduction. *Princeton university press*. 2019.
12. Eisinger, D. Spatial pattern formation facilitates eradication of infectious diseases. *The Journal of applied ecology*. 2008. T. 45. № 2. 415 с.
13. Khadilkar, H. Optimising lockdown policies for epidemic control using reinforcement learning. *Transactions of the Indian National Academy of Engineering*. 2020. T. 5. № 2. С. 129–132.
14. Ohi Abu Quwsar, MF Mridha, Monowar Muhammad Mostafa, Hamid Md Abdul Exploring optimal control of epidemic spread using reinforcement learning. *Scientific Reports (Nature Publisher Group)*. 2020. T. 10. № 1.
15. Sutton, R. S. Reinforcement learning: An introduction. *MIT press*, 2018.
16. Glorot, X. Deep sparse rectifier neural networks. *Proceedings of the fourteenth international conference on artificial intelligence and statistics / JMLR Workshop and Conference Proceedings*. 2011. С. 315–323.
17. Ioffe, S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. 2015. URL: <https://arxiv.org/abs/1502.03167>.
18. Mnih V., Kavukcuoglu K. Silver D. Human-level control through deep reinforcement learning. *Nature*. 2015. T. 518. № 7540. С. 529–533.
19. Kingma, D. P. Adam: A method for stochastic optimization. 2017. URL: <https://arxiv.org/abs/1412.6980>.
20. Баклан І.В. Класифікація моделей марковського типу: Наукова монографія // Баклан І.В., Степанова Г.А. К. : НАУ, 2012. 84 с.

Sarnatskyi V.V., Baklan I.V. THE INFLUENCE OF SOCIODEMOGRAPHIC HETEROGENEITY ON THE OPTIMAL LOCKDOWN POLICY

This paper describes the results of the study of optimal lockdown policy in simulated environments with different sociodemographic parametrization.

As part of this work, an environment for modeling the spread of the epidemic in a particular country was implemented. This model can be flexibly parameterized by setting such parameters as: the number of residents of the country, their age distribution, the distribution of the size of families, the number of educational and work institutions, the distribution of each resident's belonging to an institution of study and work, depending on age. With the help of this parameterization, models of some European countries were built. The spread of the epidemic is modeled by the contact tracing method. This method consists of simulating the movements of each resident of the country over time and his contacts with other residents. Although this modelling approach of the spread of epidemic is more demanding on computing resources, it is capable of conveying complex stochastic processes of the real world.

Based on the developed model, an agent learning environment was built, formulated as a Markov decision-making process. Three actions were chosen as the actions available to the agent: complete quarantine in all establishments, in which all residents are in contact only with their relatives; canceling quarantine and

inaction. The reward for the chosen action was formed of two parts: reward for the economic state and reward for the epidemiological state. As a condition, the agent monitors the percentage of residents infected over the past few days.

The agent's decision-making model is built on the basis of a fully-connected artificial neural network and takes the current state of the environment as input. The neural network weights are optimized using Deep Q-learning algorithm. For validation of trained agent, its effectiveness was compared to the effectiveness of the other three strategies: the full quarantine strategy, the ignore strategy, and the random action strategy. Also, agents who have been trained in environments parameterized by one country have been tested on environments based on other countries. The results of the comparison showed that in most cases an agent trained in a certain country shows more effective policy than other agents. From this we can conclude that sociodemographic distributions play an important role in the process of finding an optimal lockdown policy.

Key words: *reinforcement learning, intelligent agent, artificial intelligence, epidemiology.*