

МІНІСТЕРСТВО ОСВІТУ І НАУКИ УКРАЇНИ  
ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ ПОЛІТЕХНІЧНИЙ УНІВЕРСИТЕТ

ІНСТИТУТ КОМП'ЮТЕРНИХ СИСТЕМ

МАТЕРІАЛИ ДЕВ'ЯТОЇ  
МІЖНАРОДНОЇ НАУКОВОЇ КОНФЕРЕНЦІЇ  
СТУДЕНТІВ ТА МОЛОДИХ ВЧЕНІХ



ПРИСВЯЧЕНА 55-РІЧЧЮ  
ІНСТИТУТУ КОМП'ЮТЕРНИХ СИСТЕМ

“Сучасні інформаційні технології 2019”

“Modern Information Technology 2019”



**NetCracker®**



23-24 травня

Одеса  
«Екологія»  
2019

УДК 004.42

**НАВЧАННЯ СТРАТЕГІЧНОМУ МИСЛЕННЮ В ІГРАХ ЗА ДОПОМОГОЮ  
МАШИННОГО НАВЧАННЯ**

Стінський В. В., Комлева Н. О.

к.т.н., доцент каф. СПЗ Комлева Н. О.

Одеський Національний Політехнічний Університет, УКРАЇНА

**АНОТАЦІЯ.** Розглянуто проблематику використання ботів для навчання стратегічному мисленню. Показано переваги використання машинного навчання при реалізації подібних ботів. Описана стратегія навчання бота і стратегія навчання ботом гравця. Виконано аналіз та вибір інструментів розробки.

**Вступ.** Не існує загальноприйнятого визначення стратегічного мислення, немає загальної згоди щодо його ролі і значення, а також стандартизованого переліку ключових компетенцій стратегічних мислителів. Стратегічне мислення в іграх ніколи не виходило з моди, але йому досить важко навчитися. В різних стратегічних іграх поріг входження на вищий рівень складності являється доволі високим, окрім цього далеко не завжди у новачка є можливість тренуватися з подібними йому по рівню гри, що й створює більш всього труднощів в процесі навчання, а також, при досягненні певного високого рівня гри інколи проблематично знайти суперників, які б змусили розвиватись надалі. Рішення цієї проблеми – бот–стратег, який підлаштовується під рівень гри гравця та не дає йому робити одні й ті ж самі помилки та використовувати незмінну стратегію шляхом адаптації під стиль гравця, тим самим пришвидшуючи та полегшуючи навчання гравця.

**Мета роботи.** Метою роботи є підвищення швидкості навчання гравця стратегічному мисленню за допомогою бота, за рахунок вчення на власних помилках та плавного підвищення рівня складності гри.

**Основна частина роботи.** Під час розробки було використано такі інструменти: JavaFX – платформа, яка допомагає зручно й швидко створювати інтерфейс користувача [1]; JavaWeka – бібліотека, яка призначена для полегшення роботи із машинним навчанням у мові програмування Java [2].

Розглянемо процес навчання бота. Бот–стратег навчається на прикладі програми із штучним інтелектом AlphaGo [3] від Google, яка в 2016 році перемогла одного із найкращих гравців світу в «Го» [4] Лі Седола.

Відповідний алгоритм є наступним. З початку бот–стратег має три рівня складності: низький, середній, високий. Після 5 ігор бот визначає рівень гравця та вносить особливості його стратегії в свою систему даних, котра базується на основі дерева, де кожен вузол – це окрема ситуація на ігровій дошці. Вузол має такі характеристики: **N** (скільки разів гравець був в даному вузлі), **V** (вага цього вузла), **P** (ймовірність, що із усіх допустимих вузлів на цьому ході гравець обере саме цей).

Після визначення рівня гри гравця бот тренується сам із собою за наступним алгоритмом (рисунок 1):

1. Бере дерево, коренем якого є поточний вузол.
2. Йде в той дочірній вузол, де більше  $Q + U$  ( $U$  — добавка, стимулююча пошук нових шляхів; її значення більше на початку тренування та менше – в подальшому).
3. Таким чином, доходить до кінця дерева – стану, коли дочірніх вузлів немає, а гра все ще не закінчена.
4. Передає цей стан на вхід нейромережі, у відповідь отримує  $V$  (value, або вага поточного вузла) і  $p$  (ймовірність наступних ходів).
5. Записує  $V$  в вузол.
6. Створює дочірні вузли з  $P$  згідно  $p$  та нульовими  $N$ ,  $V$  і  $Q$ .

7. Поновлює всі ноди вище поточної, котрі були вибрані під час симуляції, таким чином:  
 $N = N + 1$ ;  $V := V + v$ ;  $Q := V / N$ .

Повторює цикл 1–7 1600 разів.

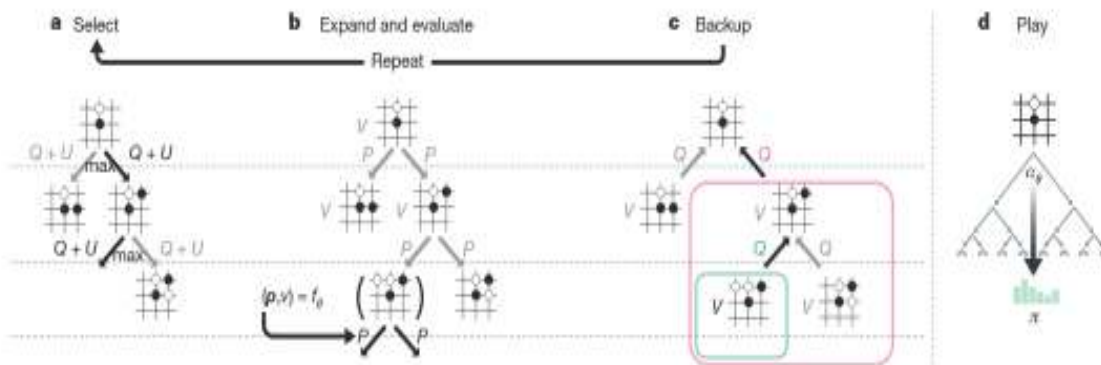


Рис. 1 – Візуалізація алгоритму тренування бота

Для передбачення  $P$  та  $V$  використовується формула

$$l = (z - v)^2 - \pi^T \log P + C \|\theta\|^2$$

Формула складається із трьох частин.

Перша частина  $(z - v)^2$  відповідає тому, що мережа повинна вміти передбачити результат, тобто  $z$  (з яким результатом закінчиться партія) яка не повинна відрізнятись від  $v$  (того value, яке вона передбачила).

В другій частині  $-\pi^T \log P$  для  $P$  використовуються наші покращенні ймовірності. Це схоже на reward в supervised learning's – потрібно як можна точніше передбачити ті ймовірності, які отримаємо після проходження по дереву; це є аналогом cross-entropy loss.

Третя частина  $C \|\theta\|^2$  – певна константа, що є регулятором.

Після навчання при виборі ходу бот використовує наступну формулу:

$$v_i = C \times \sqrt{\frac{\ln N}{n_i}}$$

де  $v$  – value вузла,  $n$  – скільки разів гравець був в цьому вузлі,  $N$  – скільки разів гравець був в батьківському вузлі,  $C$  – коефіцієнт, який впливає на важкість гри.

Розглянемо процес навчання гравця ботом. Після адаптації під стиль гравця бот грає по тактиці, яка протидіє основним елементам стратегії гравця, що сильно не завищує рівень важкості, а також адаптується під раптову зміну стиля гри гравця. Після однієї або декількох таких ігор гравець помічає основні слабкі сторони в своїй стратегії та починає її покращувати, не потрапляючи в тупикові чи повністю програшні ситуації, в які він потрапляв до цього.

**Висновок.** В процесі навчання, який тривав неділю, даним методом у гравця з початковими знаннями збільшилась швидкість розвитку стратегічного мислення в 2 рази. Покращились показники аналізу своєї стратегії та стратегії супротивника в 4 рази. Розвинулось вміння прорахувати ходи супротивника і виявляти в його стратегії слабкі місця. В грі з реальними супротивниками гравець відразу вийшов на середній рівень гри, мінуючи початковий та не допускаючи простих помилок.

### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. DiMarzio J.F. Quick Start Guide to JavaFX. McGraw-Hill Education, 2014. – 264 p.
2. Getting Started With Weka 3 – Machine Learning on GUI. URL: <https://towardsdatascience.com/getting-started-with-weka-3-machine-learning-on-gui-7c58ab684513> – Назва з екрану.
3. JustRoo. AlphaGo Zero совсем на пальцах. 2017. URL: <https://habr.com/ru/post/343590/> – Назва з екрану.
4. Саїтов Т., Гришин І. Опис гри го. URL: <http://www.go-hobby.ru/> – Назва з екрану.