Національний технічний університет України
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

MMCA

Department of Mathematical Methods of System Analysis

Reinforcement learning

Work program of the discipline (Syllabus)

Details of the discipline	
Level of higher education	***First (educational)***
Field of knowledge	*12 Information Technology*
Speciality	*122 Computer Science, 124 System Analysis*
Educational program	*System Analysis*
Discipline status	*Custom*
Form of study	*full-time (daytime)/full-time (evening)/part-time/remote/mixed*
Year of preparation, semester	*4 year, spring semester*
Scope of discipline	*3.5 ECTS credits*
Semester control / control measures	*Passed*
Timetable	*Asynchronously* *Zoom Meeting ID: 918 3662 5010 Passcode: 651168*
Language of instruction	*Ukrainian/English*
Information about Course Leader / Instructors	Lecturer: *Doctor of Physical and Mathematical Sciences, Professor, Corresponding Member of the National Academy of Sciences of Ukraine Pavlo Kasyanov Olegovich, kasyanov.pavlo@lll.kpi.ua https://www.facebook.com/pkasyanov https://www.linkedin.com/in/pavlokasyanov/ https://www.researchgate.net/profile/Pavlo_Kasyanov* Practical / Seminar: PhD student, Senior Research Engineer at SQUAD Andrey Nikolaevich Titarenko, Doctor of Physical and Mathematical Sciences, Professor, *Corresponding Member of the National Academy of Sciences of Ukraine* Pavlo Kasyanov Olegovich, kasyanov.pavlo@lll.kpi.ua https://www.facebook.com/pkasyanov https://www.linkedin.com/in/pavlokasyanov/ https://www.researchgate.net/profile/Pavlo_Kasyanov
Course Placement	https://piazza.com/national_technical_university_of_ukraine_igor_sikorsky_kyiv_polytechnic_institute/spring2024/ka0x/home

The program of the discipline

1. Description of the discipline, its purpose, subject of study and learning outcomes

The purpose *of the credit module is to form students' systematic scientific worldview, general cultural outlook and competencies to identify, pose and solve research problems in the field of computer science, evaluate and ensure the quality of research performed. In particular, to master both the fundamental principles of the theory of step-by-step decision-making (the theory of Markov decision-making processes) and dynamic programming, and to be able to apply the obtained theoretical knowledge for solving applied, in particular, problems of optimal decision-making in industry (technical support of industrial systems, industrial safety examination system); robotics (automated forecasting); business (marketing, inventory management); computer science (troubleshooting networks, optimizing requests to distributed database servers); state security and military sciences (search for moving targets, target identification, distribution of weapons); health care (medical diagnostics, development of treatment protocols). Students must master the following* ***competencies****:*

general *- GC 1 Ability to apply knowledge in practical situations; GC 3 Ability to think abstractly, apply methods of analysis and synthesis; GC 7Ability to search, process and analyze information from various sources; GC 11 Ability to generate new ideas (creativity); GC 12 Ability to work in a team and autonomously execute team decisions;*

professional *– FC 1 Ability to use system analysis as a modern interdisciplinary methodology based on examples of mathematical methods and modern information technologies, and focused on solving problems of analysis and synthesis of technical, economic, social, environmental and other complex systems; FC 6 Ability to computer implementation of mathematical models of real systems and processes; design, apply and maintain simulation software, decision-making, optimization, information processing, data mining; FC 7 Ability to use modern information technologies for computer implementation of mathematical models and forecasting the behavior of specific systems, namely: object-oriented approach in the design of complex systems of various nature, applied mathematical packages, the use of databases and knowledge; FC 10 Ability to design experimental and observational studies and analyze the data obtained from them.*

Upon completion of the course, students should ***acquire the following program learning outcomes:*** *PRN 9 Be able to create effective algorithms for computational problems of system analysis and decision support systems; PRN 12 Apply methods and means of working with data and knowledge, methods of mathematical, logical-semantic, object and simulation modeling, technologies of system and static analysis; PRN 14 Understand and apply in practice the methods of static modeling and forecasting, evaluate the initial data; PRN 17 Preserve and multiply the achievements and values of society based on an understanding of the place of the subject area in the general system of knowledge.*

Subject of study.

Tasks and classes of reinforcement learning methods are just like the area of knowledge that includes the tasks of step-by-step optimal decision-making with partial observations

The main tasks of the credit module.

According to the requirements of the program of the discipline, postgraduate students after mastering the credit module must demonstrate the following learning outcomes:

Knowledge*:*

methods and means of reinforcement learning.

Skills*:*

solve real-world problems using reinforcement learning methods and algorithms.

In particular, to formalize the problem of step-by-step optimal decision-making as a partially observable Markov decision-making process with possibly unknown transient probabilities and rewards, to apply modern algorithms for approximate solution of such problems, the ability to use relevant information technologies and create their own software products to solve real problems making optimal decisions in industry (technical support of industrial systems, industrial safety examination system); robotics (automated forecasting); business (marketing, inventory management); computer science (troubleshooting networks, optimizing requests to distributed database servers); state security and military sciences (search for moving targets, target identification, distribution of weapons); health care (medical diagnostics, development of treatment protocols).

Experience*:*

creation of a research laboratory for reinforcement learning (a paradigm of organized collaboration based on the experience of leading national laboratories in the United States), where the role of each team

member is to specialize in a particular task in order to become the best at it, while having a holistic view of the entire process.

2. **Prerequisites and post-requisites of the discipline (place in the structural and logical scheme of training in the relevant educational program)**

Basic level of English, higher mathematics, FP, OOP.

3. **The content of the discipline**

Credit module 1.

1. *Markov Decision-Making Processes*
2. *Q-Learning for Tabular Problems*
3. *Q-Approximation-Based Learning for Reinforcement Deep Learning Tasks*
4. *Approximate Dynamic Programming*
5. *Policy gradient methods*
6. *Actor-critic methods*
7. *Approximate Deep Learning with Reinforcement*

Recommended topics of practical (seminar) classes

The purpose of conducting practical classes is to consolidate the knowledge gained in lectures, to acquire the ability to solve real problems of step-by-step optimal decision-making using methods and means of reinforcement learning.

1. *Introductory lesson. Downloading useful resources.*
2. *The Task of the Multi-Armed Bandit*
3. *Markov decision-making processes. Dynamic programming methods. Bellman's optimality equation.*
4. *Monte Carlo methods*
5. *Time Difference Method*
6. *Sarsa, Expected Sarsa, Dyna-Q, Q-learning algorithms,*
7. *Tile coding, Keras and TensorFlow libraries for reinforcement deep learning tasks,*
8. *Gradient and semi-gradient methods,*
9. *Gaussian Actor-Critic Method*

4. **Training Materials & Resources**

All the necessary materials are contained on the Piazza platform

https://piazza.com/national_technical_university_of_ukraine_igor_sikorsky_kyiv_polytechnic_institute/spring2024/ka0x/home

Basic literature:

1. Reinforcement Learning, second edition: An Introduction (Adaptive Computation and Machine Learning series): Sutton, Richard S., Barto, Andrew G.: 9780262039246: Amazon.com: Books
2. (PDF) Algorithms for reinforcement learning | Csaba Szepesvari - Academia.edu
3. Markov Decision Processes | Wiley Series in Probability and Statistics
4. ELAKPI: System Analysis of Stochastic Distributed Systems
5. https://www.coursera.org/specializations/reinforcement-learning
6. https://piazza.com/national_technical_university_of_ukraine_igor_sikorsky_kyiv_polytechnic_institute/spring2024/ka0x/home

Further reading:

https://drive.google.com/drive/folders/1V9jAShWpccLvByv5S1DuOzo6GVvzd4LV

Educational content

5. Methods of mastering the discipline (educational component)

5.1. Lectures

Salary No.	*Title of the topic of the lecture and a list of the main questions (list of didactic aids, references to literature and tasks for the SRS)*
1	*Markov decision-making processes. [1-6] (Year 6)*
2	*Q-learning for tabular problems. [1-6] (Year 2)*
3	*Q-learning based on approximations for deep reinforcement learning tasks. [1-6] (Year 2)*
4	*Approximate dynamic programming. [1-6] (Year 2)*
5	*Policy gradient methods. [1-6] (Year 2)*
6	*Actor-critic methods. [1-6] (Year 2)*
7	*Approximate Deep Reinforcement Learning. [1-6] (Year 2)*

5.2. Practical exercises

The purpose of conducting practical classes is to consolidate the knowledge gained in lectures, to acquire the ability to solve real problems with the help of financial analytical simulations

Salary No.	*Name of the topic of the lesson (list of didactic support, links to literature and tasks for the SRS)*
1	*Introductory lesson. Downloading useful resources. [4-6] (2 hours)*
2	*The task of a multi-armed bandit. [4-6] (2 hours)*
3	*Markov decision-making processes. Dynamic programming methods. Bellman's optimality equation. [4-6] (2 hours)*
4	*Monte Carlo Methods [4-6] (2 hours)*
5	*Time difference method. [4-6] (2 hours)*
6	*Sarsa, Expected Sarsa, Dyna-Q, Q-learning algorithms. [4-6] (2 hours)*
7	*Tile coding, Keras and TensorFlow libraries for reinforcement deep learning tasks. [4-6] (2 hours)*
8	*Gradient and semi-gradient methods. [4-6] (2 hours)*
9	*The Gaussian Actor-Critic Method [4-6] (2 hours)*

6. Independent work of a student/graduate student

Students' independent work consists in processing materials and completing tasks on the Piazza distance learning platform

https://piazza.com/national_technical_university_of_ukraine_igor_sikorsky_kyiv_polytechnic_institute/spring2024/ka0x/home

Policy & Control

7. Academic discipline policy (educational component)

Proper completion of all tasks on the Piazza distance learning platform is required https://piazza.com/national_technical_university_of_ukraine_igor_sikorsky_kyiv_polytechnic_institute/spring2024/ka0x/home *according to the requirements and individual strategy, which is determined by the trainee independently or, if necessary, under the scientific guidance of the teacher / supervisor.*

8. Types of control and rating system for assessing learning outcomes (CRO)

***Current control**: each student determines the strategy for completing tasks (independently or, if necessary, under the scientific guidance of the teacher / supervisor), aiming to receive 100 points at the end of the semester.*

***Types of control** :*

a) model tasks: 5 homework assignments on the distance learning platform **https://piazza.com/national_technical_university_of_ukraine_igor_sikorsky_kyiv_polytechnic_institute/spring2024/ka0x/home** *(there are recommended deadlines), each of which can be evaluated with a maximum of 20 points;*

b) practical problems: implementation and video presentation of a practical project as part of a team (1-5 students). A maximum of 85 points (the number of points for the project is equal to the sum of the peer review (maximum 40 points) and the number of positive reactions to the project's video (e.g., **YouTube***) and its code (e.g.,* **GitHub: Where the world builds software · GitHub***)".*

c) Educational and Methodological Problems: recording of a video lecture on a book [Reinforcement Learning, second edition: An Introduction (Adaptive Computation and Machine Learning series): Sutton, Richard S., Barto, Andrew G.: 9780262039246: Amazon.com: Books] as part of a team (1-5 students), each of whom can be valued at 75 points by voting (positive reactions, comments, etc.) of other team members (e.g., YouTube).

G) Incentive points for completing tasks to improve didactic materials In disciplines, from 20 to 40 incentive points are provided.

***Calendar control**: it is carried out as a monitoring of the current state of fulfillment of the requirements of the syllabus (in proportion to the number of working weeks per semester).*

***Semester control**: differentiated credit (sum of points for the semester, additional performance of types of work items a, b, c, d)*

***Conditions for admission to semester control**: it is desirable to have a semester rating of at least 20 points.*

Table of correspondence of rating points to grades on the university scale:

Score	Score
100-95	Perfectly
94-85	Very good
84-75	Well
74-65	Satisfactory
64-60	Enough
Less than 60	Disappointing
Admission conditions are not met	Not allowed

9. Additional information on the discipline (educational component)

All the necessary materials are contained on the Piazza learning platform
https://piazza.com/national_technical_university_of_ukraine_igor_sikorsky_kyiv_polytechnic_institute/spring2024/ka0x/home

Work program of the discipline (syllabus):

Compiled by Director of IASA, Doctor of Physical and Mathematical Sciences, Professor, Corresponding Member of the National Academy of Sciences of Ukraine Kasyanov Pavlo Olegovych

Ph.D., Leading Research Engineer of SQUAD, Andrey Nikolaevich Titarenko

Approved by the Department of Mathematical Methods of System Analysis (Minutes No. 13 dated 05.06.2024)

Approved by the Methodological Commission of the Faculty (Minutes No. 10 dated 24.06.2024)