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	***Third (educational and scientific)***
Field of knowledge	*12 Information Technology*
Speciality	*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	*2nd year, autumn/spring semester*
Scope of discipline	*4 ECTS credits*
Semester control / control measures	*Passed*
Timetable	
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 Kasyanov Pavlo Olehovych, 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: *Doctor of Physical and Mathematical Sciences, Professor, Corresponding Member of the National Academy of Sciences of Ukraine Kasyanov Pavlo Olehovych*, 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	KA 1X | Class Profile | Piazza

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 in applicants for the third level of higher education (PhD) the abilities of a 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 learn both the fundamental principles of the theory of step-by-step decision-making (the theory of Markov decision-making processes) and dynamic programming, and be able to apply the acquired theoretical knowledge to solve applied, in particular, problems of 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), as well as postgraduate students must master the following* ***competencies****:*

general *- GC 4 Ability to independently conduct research activities, including analysis of problems, setting goals and objectives, selection of means and methods of research, as well as assessment of its quality; GC*

5 Ability to initiate, plan, implement and adjust a sequential process of thorough scientific research; GC 6 Ability to critically analyze, evaluate, and synthesize new and complex ideas; GC 7 Ability for continuous self-development and self-improvement;

professional *– FC 1 Ability to initiate complex projects using a systematic approach and implement them independently; FC 2 Ability to comply with moral and ethical rules of conduct, research ethics, characteristics for participants in the academic environment, as well as the rules of academic integrity in scientific research; FC 3 Ability to critically analyze the positive and negative qualities of existing methods of system analysis, as well as to assess their capabilities for further use in solving specific scientific and practical problems; FC 4 Ability to make scientifically sound decisions in conditions of uncertainty, which requires the development of new methods and the conduct of research and innovation activities; FC 5 Ability to carry out research and professional activities at an interdisciplinary level; FC 6 Ability to deeply analyze and create new methods for analyzing data and knowledge; FC 7 Ability to perform research on loosely structured problems, develop new methods, and then solve them; FC 8 Ability to plan and conduct scientific research, prepare, present and publish the results of research activities.*

Upon completion of the course, applicants for the third level of higher education should ***acquire the following program learning outcomes:*** *PRN 4 Know the advantages and disadvantages of existing methods of system analysis and the possibility of their use to solve specific scientific and applied problems in intelligent decision support systems; PRN 5 Know the basics of the organization of the research scientific process to solve significant problems in the field of system analysis, be able to apply knowledge of the basics of analysis and synthesis in various subject areas, critical comprehension and solution of research problems; PRN 10 Be able to create new methods of system analysis and mathematical models of complex systems of various nature; PRN 11 Be able to develop and use new methods for analyzing complex systems and new methods of decision-making under uncertainty; PRN 12 Be able to critically analyze the advantages and disadvantages of known methods of system analysis, as well as be able to assess the possibilities of their use to solve specific scientific and practical problems; PRN 13 Be able to develop scientific projects in the field of system analysis; PRN 14 Be able to implement the results of scientific research based on the methods of system analysis; PRN 15 Be able to solve complex problems in the field of system analysis or as a result of research and innovation activities, which involves a deep rethinking of existing and the creation of new holistic knowledge; PRN 17 Read and understand foreign texts in the specialty; freely present and discuss with specialists and non-specialists the results of research, scientific and applied problems of the industry in the state and foreign languages, competently reflect the results of research in scientific publications in leading international scientific journals; PRN 18 Adhere to the rules of academic integrity; know and adhere to the basic principles of academic integrity in scientific and educational (pedagogical) activities.*

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, programming.

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

KA 1X | Class Profile | Piazza

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. CSCPHD | Class Profile | Piazza

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
KA 1X | Class Profile | Piazza*; preparation for the test.*

Policy & Control

7. Academic discipline policy (educational component)

Proper completion of all tasks on the Piazza distance learning platform is required
KA 1X | Class Profile | Piazza *according to the requirements and individual strategy, which is determined by the graduate student 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:
1) two answers (each student on average) in practical classes (provided that an average of 8 students are interviewed in one lesson with a maximum group size of 30 people);
2) performance of one test (remotely – tests and tasks).

RATING POINTS SYSTEM

1. Practical lesson
The maximum number of points in all practical classes is 20 points X 2 = 40 points.
Evaluation criteria:
0-8 points – the problem is not solved, while the student has certain theoretical information about the topic of the practical lesson;
9-14 points – the problem is not fully solved or the solution contains gross technical shortcomings, while the student is fluent in theoretical information about the topic of the practical lesson;
15-20 points – the problem is solved as a whole, while the student is fluent in theoretical information about the topic of the practical lesson.

2. Modular control.
The maximum number of points for a test (project) is 60 points.
Evaluation criteria:
0-20 points – the problem as a whole has not been solved or the solution contains gross technical shortcomings, there is no answer to the theoretical question;
21-50 points – the problem is solved as a whole, the theoretical issue is disclosed;
51-60 points – the problem is solved, the answer to the theoretical question is exhaustive.

Penalty and incentive points for:

- performance of tasks to improve didactic materials in disciplines is given from 15 to 30 incentive points.

According to the results of academic work in the first 7 weeks, the "ideal student" should score 20 points. At the first attestation (8th week), a student receives an "enrolled" if his current rating is not less than 10 points. According to the results of 13 weeks of study, the "ideal student" should score 40 points. At the second attestation (14th week), the student receives "passed" if his current rating is not less than 20 points. The maximum amount of points is 100. A prerequisite for admission to the test is a positive mark on the test. To receive credit from the credit module "automatically", you need to have a rating of at least 60 points, as well as a credited test (more than 30 points). Students who have a rating of less than 60 points at the end of the semester, as well as those who want to improve their grade in the ECTS system, complete the test work. At the same time, the points for the test are added to the points for the test work, and this rating score is final. The control task of this work consists of two questions of the work program from the list provided in the methodological recommendations for mastering the credit module. An additional question on the topics of practical classes is given to students who did not take part in the work of a particular practical lesson. An unsatisfactory answer to an additional question lowers the overall score by 4 points.
Each question is scored out of 20 points according to the grading system:
– "excellent", full answer (at least 90% of the required information) – 20... 18 points;
– "good", a fairly complete answer (at least 75% of the required information, or minor inaccuracies) – 17... 14 points;
– "satisfactory", incomplete answer (at least 60% of the required information and some errors) – 13... 11 points;
– "unsatisfactory", unsatisfactory answer – 0 points.
Sum of points: for each of the two questions of the test and the test, it is transferred to the credit grade according to the table.

Points Score
100… 95 Excellent
94… 85 Very Good
84… 75 Good
74… 65 Satisfactory
64… 60 Enough
Less than 60 Unsatisfactory
R&D Not Credited Not Allowed

9. Additional information on the discipline (educational component)

All the necessary materials are contained on the Piazza learning platform
KA 1X | Class Profile | Piazza

Work program of the discipline (syllabus):

Compiled by Director of IASA, Doctor of Physical and Mathematical Sciences, Professor, Kasyanov Pavel Olegovich

Acad. Sci. National Academy of Sciences of Ukraine, Doctor of Technical Sciences, Prof. Mikhail Zakharovich Zgurovsky

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)