Federal State Autonomous Educational Institution of Higher Education "Moscow Institute of Physics and Technology (National Research University)"

APPROVED
Head of the Phystech School of Applied Mathematics and Informatics
A.M. Raygorodskiy

Work program of the course (training module)

course:	Reinforcement Learning/Обучение с подкреплением
major:	Applied Mathematics and Informatics
specialization:	Contemporary Combinatorics/Современная комбинаторика
	“Pusk” Online and Supplementary Education Centre
	Chair of Discrete Mathematics
term:	2
qualification:	Master

Semester, form of interim assessment: 3 (fall) - Grading test

Academic hours: 45 AH in total, including:
lectures: 15 AH.
seminars: 30 AH.
laboratory practical: 0 AH.

Independent work: 45 AH.

In total: 90 AH, credits in total: 2

Author of the program: R.G. Neychev, professor

The program was discussed at the Chair of Discrete Mathematics 05.03.2020

Annotation

This course succeeds the Machine Learning course and aims to introduce students to the contemporary state of Reinforcement Learning. The biological inspiration and theoretical foundations of Reinforcement Learning algorithms are reinforced with comprehensive practical assignments. The course covers materials from genetic algorithms to the latest achievements in Reinforcement Learning approaches and narrows the gap between different sub-areas of Artificial Intelligence.

1. Study objective

Purpose of the course

- Learn how to apply Reinforcement Learning techniques in practice
- Get familiar with both fundamental and most recent approaches in Reinforcement Learning

Tasks of the course

- Reinforcement Learning problem statement and ability to develop the general pipeline of the solution
- Ability to apply the Reinforcement Learning techniques to the real world problems
- Essential experience with PyTorch framework

2. List of the planned results of the course (training module), correlated with the planned results of the mastering the educational program

Mastering the discipline is aimed at the formation of the following competencies:

Code and the name of the competence	Competency indicators
UC-1 Use a systematic approach to critically analyze a problem, and develop an action plan	UC-1.2 Search for solutions by using available sources
Gen.Pro.C-1 Address current challenges in fundamental and applied mathematics	Gen.Pro.C-1.3 Understand interdisciplinary relations in applied mathematics and computer science and apply them in professional tasks
Gen.Pro.C-2 Iimprove upon and implement new mathematical methods in applied problem solving	Gen.Pro.C-2.3 Understand professional terminology used in modern scientific and technical literature and present scientific results in oral and written form
Pro.C-1 Become part of a professional community and conduct local research under scientific guidance using methods specific to a particular professional setting	Pro.C-1.2 Understand the verification process of software models used to solve related scientific problems
Pro.C-2 Understands and is able to apply modern mathematical apparatus and algorithms, the basic laws of natural science, modern programming languages and software; operating systems and networking technologies in research and applied activities	Pro.C-2.2 Demonstrate practical experience of applying methods and digital signal processing algorithms, using the Internet, abstracting, referencing, searching for bibliographic sources, and working with scientific sources

3. List of the planned results of the course (training module)

As a result of studying the course the student should:

know:

- statement and solution of the problem of synthesizing an object with a given combinatorial characteristic;
- an approach to the analysis of cyclic computations based on a fixed point.
- statement and solution of the problem of synthesizing a data structure with specified mathematical properties;
- the relationship between different formulations of the theory of computation;
- various options for immersing object theories.
- the process of compiling combinatorial code;
- connection of syntax and semantics of calculations with selected bases;
- various mechanisms of calculations and ways of improving them by means of various parameterizations;
- ways and methods of eliminating collisions of variables;
- various forms, including equational, theory of computation;
- the cycle of the abstract machine;
- perspectives of applicative computational technologies and languages CAML, Haskell, F #

be able to:

- to synthesize and analyze an object with a given combinatorial characteristic;
- to compute (interpret) the combinatorial program code containing loop constructions;
- to establish a combinatorial basis of calculations and apply it to solve the problem of compiling a combinatorial code;
- to build equational representations of calculations;
- to carry out reduction of abstraction to supercombiners;
- to calculate (interpret) the reduced expression;
- optimize computations by applying parameterizations;
- to perform code generation of the original expression into an intermediate representation;
- optimize and execute the generated code based on the instructions of the abstract machine;
- perform calculations involving a fixed point.

master:

- have practical skills in building and applying simulation models of distributed computing.

4. Content of the course (training module), structured by topics (sections), indicating the number of allocated academic hours and types of training sessions

4.1. The sections of the course (training module) and the complexity of the types of training sessions

№	Topic (section) of the course	Types of training sessions, including independent work			
		Lectures	Seminars	Laboratory practical	Independent work
1	Reinforcement Learning problem statement.	3	6		10
2	Rewards discounting in RL.	3	6		10
3	On policy and off policy algorithms. N-step algorithms	3	6		10
4	DQN	3	6		10
5	DDQN	3	6		5
AH in total		15	30		45
Exam preparation		0 AH.			
Total complexity		90 AH., credits in total 2			

4.2. Content of the course (training module), structured by topics (sections)

Semester: 3 (Fall)

1. Reinforcement Learning problem statement.

Stochastic and black box optimization. Value based methods in RL

2. Rewards discounting in RL.

Value iteration. Policy iteration. Model free learning. Q-learning, SARSA

3. On policy and off policy algorithms. N-step algorithms

- Approximate Q-learning
- Value function approximation using complex functions and neural networks.

4. DQN

Experience replay buffer Autocorrelation problem Policy gradient for sequence modeling. Self-critical sequence training

5. DDQN

Policy gradient. REINFORCE algorithm. A2C, A3C Policy gradient as optimization approach in different areas.

5. Description of the material and technical facilities that are necessary for the implementation of the educational process of the course (training module)

A standard classroom.

6. List of the main and additional literature, that is necessary for the course (training module) mastering

Main literature

1. Функциональное и логическое программирование [Текст] : учеб. пособие для вузов / А. Л. Ездаков .— М. : БИНОМ. Лаб. знаний, 2009 .— 119 с.
2. Обучение с подкреплением [Электронный ресурс] / Р. С. Саттон, Э. Г. Барто ; пер. с англ. Е. О. Романова под ред. Ю. В. Тюменцева .— 2-е изд. — М. : Бином. Лаборатория знаний, 2014 .— (Адаптивные и интеллектуальные системы) .— Электрон. версия печ. публикации .— Полный текст (Доступ из сети МФТИ).

Additional literature

1. Машинное обучение [Текст] = Real-World Machine Learning / Х. Бринк, Дж. Ричардс, М. Феверолф .— СПб. : Питер, 2017 .— 336 с.: ил. — (Библиотека программиста). - 1000 экз. - ISBN 978-5-496-02989-6.) .— Полный текст (Доступ из сети МФТИ / Удаленный доступ).

7. List of web resources that are necessary for the course (training module) mastering

http://dm.fizteh.ru/

8. List of information technologies used for implementation of the educational process, including a list of software and information reference systems (if necessary)

Multimedia technologies can be employed during lectures and practical lessons, including presentations.

9. Guidelines for students to master the course

1. Statement of the reinforcement learning problem.
2. Value-based methods
3. Model-free methods
4. Approximate RL and Deep Q-learning
5. Exploration in reinforcement learning
6. Policy-based methods

SUPPLEMENT

Assessment funds for course (training module)

major:	Applied Mathematics and Informatics
specialization:	Contemporary Combinatorics/Современная комбинаторика
	"Pusk" Online and Supplementary Education Centre
	Chair of Discrete Mathematics
term:	2
qualification:	Master

Semester, form of interim assessment: 3 (fall) - Grading test

Author: R.G. Neychev, professor

1. Competencies formed during the process of studying the course

Code and the name of the competence	Competency indicators
UC-1 Use a systematic approach to critically analyze a problem, and develop an action plan	UC-1.2 Search for solutions by using available sources
Gen.Pro.C-1 Address current challenges in fundamental and applied mathematics	Gen.Pro.C-1.3 Understand interdisciplinary relations in applied mathematics and computer science and apply them in professional tasks
Gen.Pro.C-2 Iimprove upon and implement new mathematical methods in applied problem solving	Gen.Pro.C-2.3 Understand professional terminology used in modern scientific and technical literature and present scientific results in oral and written form
Pro.C-1 Become part of a professional community and conduct local research under scientific guidance using methods specific to a particular professional setting	Pro.C-1.2 Understand the verification process of software models used to solve related scientific problems
Pro.C-2 Understands and is able to apply modern mathematical apparatus and algorithms, the basic laws of natural science, modern programming languages and software; operating systems and networking technologies in research and applied activities	Pro.C-2.2 Demonstrate practical experience of applying methods and digital signal processing algorithms, using the Internet, abstracting, referencing, searching for bibliographic sources, and working with scientific sources

2. Competency assessment indicators

As a result of studying the course the student should:

know:

- statement and solution of the problem of synthesizing an object with a given combinatorial characteristic;
- an approach to the analysis of cyclic computations based on a fixed point.
- statement and solution of the problem of synthesizing a data structure with specified mathematical properties;
- the relationship between different formulations of the theory of computation;
- various options for immersing object theories.
- the process of compiling combinatorial code;
- connection of syntax and semantics of calculations with selected bases;
- various mechanisms of calculations and ways of improving them by means of various parameterizations;
- ways and methods of eliminating collisions of variables;
- various forms, including equational, theory of computation;
- the cycle of the abstract machine;
- perspectives of applicative computational technologies and languages CAML, Haskell, F #

be able to:

- to synthesize and analyze an object with a given combinatorial characteristic;
- to compute (interpret) the combinatorial program code containing loop constructions;
- to establish a combinatorial basis of calculations and apply it to solve the problem of compiling a combinatorial code;
- to build equational representations of calculations;
- to carry out reduction of abstraction to supercombiners;
- to calculate (interpret) the reduced expression;
- optimize computations by applying parameterizations;
- to perform code generation of the original expression into an intermediate representation;
- optimize and execute the generated code based on the instructions of the abstract machine;
- perform calculations involving a fixed point.

master:

- have practical skills in building and applying simulation models of distributed computing.

3. List of typical control tasks used to evaluate knowledge and skills

1. RL problem statement. State, Action, Reward, Environment, Action definitions.
2. Crossentropy method
3. Value function, Q-function
4. Q-learning, approximate Q-learning. DQN, bells and whistles (Experience replay, Double DQN, autocorrelation problem)
5. Policy gradient and REINFORCE algorithm
6. Baseline in policy gradient
7. A2C
8. Policy gradient applications in other domains (outside RL). How Self-Critical Sequence Training is performed? What is used as a baseline?

4. Evaluation criteria

Questions for the differentiated credit

1. Prove that if m, n are two coprime integers of different parity, then the numbers m2 - n2 and 2mn are also coprime.
2. Write and prove the general formula for the number of different representations of a given integer n as the sum of two squares. (Representatives that are not obtained from each other by changing signs and the order of the terms are considered different.)
3. Based on the obtained formula, derive the lower bound for the maximum number of equal distances among the given n points on the plane using a regular rectangular lattice.
4. Build a regular pentagon using a compass and a ruler.
5. Build a regular 15-gon using a compass and a ruler.
6. You are given a single segment. It is required to construct using a compass and a ruler a segment of length x satisfying the equation
7. Based on the previous task, prove that a regular heptagon cannot be built using a compass and a ruler.
8. Prove that trisection of the angle is impossible.
9. Describe all possible combinations of the amounts of black and white balls in the ballot box, so that if two balls are randomly fished in a sample without returning, the probability of fishing two white balls is exactly 0.5.
10. Consider the relation on the sides a, b, c of the triangle, in which a triangle with vertices at the bases of the bisectors is isosceles. Assuming that the sides converging on side c of the large triangle are equal, reduce this relation to the following
11. In what follows, we consider the cube defined by the first of the three equations (refusing the requirement that a, b, c be sides of a triangle). Show that the resulting cube is indecomposable, that is, the polynomial that defines it does not factor.
12. In addition to this, show that our cube is nonsingular, that is, there is not a single point on its projectivization at which each direction is tangent (or the same thing at which all three first partial derivatives of the polynomial defining it degenerate).

Assessment “excellent (10)” is given to a student who has displayed comprehensive, systematic and deep knowledge of the educational program material, has independently performed all the tasks stipulated by the program, has deeply studied the basic and additional literature recommended by the program, has been actively working in the classroom, and understands the basic scientific concepts on studied discipline, who showed creativity and scientific approach in understanding and presenting educational program material, whose answer is characterized by using rich and adequate terms, and by the consistent and logical presentation of the material;

Assessment “excellent (9)” is given to a student who has displayed comprehensive, systematic knowledge of the educational program material, has independently performed all the tasks provided by the program, has deeply mastered the basic literature and is familiar with the additional literature recommended by the program, has been actively working in the classroom, has shown the systematic nature of knowledge on discipline sufficient for further study, as well as the ability to amplify it on one’s own, whose answer is distinguished by the accuracy of the terms used, and the presentation of the material in it is consistent and logical;

Assessment “excellent (8)” is given to a student who has displayed complete knowledge of the educational program material, does not allow significant inaccuracies in his answer, has independently performed all the tasks stipulated by the program, studied the basic literature recommended by the program, worked actively in the classroom, showed systematic character of his knowledge of the discipline, which is sufficient for further study, as well as the ability to amplify it on his own;

Assessment “good (7)” is given to a student who has displayed a sufficiently complete knowledge of the educational program material, does not allow significant inaccuracies in the answer, has independently performed all the tasks provided by the program, studied the basic literature recommended by the program, worked actively in the classroom, showed systematic character of his knowledge of the discipline, which is sufficient for further study, as well as the ability to amplify it on his own;

Assessment “good (6)” is given to a student who has displayed a sufficiently complete knowledge of the educational program material, does not allow significant inaccuracies in his answer, has independently carried out the main tasks stipulated by the program, studied the basic literature recommended by the program, showed systematic character of his knowledge of the discipline, which is sufficient for further study;

Assessment “good (5)” is given to a student who has displayed knowledge of the basic educational program material in the amount necessary for further study and future work in the profession, who while not being sufficiently active in the classroom, has nevertheless independently carried out the main tasks stipulated by the program, mastered the basic literature recommended by the program, made some errors in their implementation and in his answer during the test, but has the necessary knowledge for correcting these errors by himself;

Assessment “satisfactory (4)” is given to a student who has discovered knowledge of the basic educational program material in the amount necessary for further study and future work in the profession, who while not being sufficiently active in the classroom, has nevertheless independently carried out the main tasks stipulated by the program, learned the main literature but allowed some errors in their implementation and in his answer during the test, but has the necessary knowledge for correcting these errors under the guidance of a teacher;

Assessment “satisfactory (3)” is given to a student who has displayed knowledge of the basic educational program material in the amount necessary for further study and future work in the profession, not showed activity in the classroom, independently fulfilled the main tasks envisaged by the program, but allowed errors in their implementation and in the answer during the test, but possessing necessary knowledge for elimination under the guidance of the teacher of the most essential errors;

Assessment “unsatisfactory (2)” is given to a student who showed gaps in knowledge or lack of knowledge on a significant part of the basic educational program material, who has not performed independently the main tasks demanded by the program, made fundamental errors in the fulfillment of the tasks stipulated by the program, who is not able to continue his studies or start professional activities without additional training in the discipline in question;

Assessment “unsatisfactory (1)” is given to a student when there is no answer (refusal to answer), or when the submitted answer does not correspond at all to the essence of the questions contained in the task.

5. Methodological materials defining the procedures for the assessment of knowledge, skills, abilities and/or experience

During differentiated credit the student are allowed to use the program of the discipline.