Syllabus
Course Program

Graph Data Analysis

Specialty
113 Applied mathematics

Institute
Educational and Scientific Institute of Computer Science and Information Technology

Educational program
Intelligent Data Analysis

Department
Computer Mathematics and Data Analysis (324)

Level of education
Master's level

Discipline type
Special (professional), Selective

Semester
1

Language of instruction
Ukranian, English

Lecturers and course developers

Vladyslav Kolbasin

vladyslav.kolbasin@khpi.edu.ua
Senior teacher of the Department of Computer Mathematics and Data Analysis

Work experience - 15 years. Leading lecturer in the disciplines: "Object-oriented programming", "Computer geometry and graphics"

More about the lecturer on the website

General information

Summary

The discipline is aimed at mastering the theoretical and practical basics of working with graph data based on the use of artificial intelligence models. The course covers the tasks of building embeddings of nodes, connections, and graphs; as well as their use for solving classification problems.

Course objectives and goals

The goal of the discipline is the formation of ideas about the main stages, tasks, and methods of solving problems that arise during the analysis of large graphs.

Format of classes

Lectures, laboratory works, self-study, consultations. Final control in the form of an exam.

Competencies

GC 3. The ability to continuously study, acquire new knowledge and skills, including in a field other than a professional one.
GC 4. The ability to identify, pose and solve problems in professional activities.
GC 6. The ability to critically evaluate and rethink the accumulated experience (own and others), to analyze one's professional and social activities.
GC 8. The ability to effectively build communication, based on the goals and situation of communication.

SC 1. The ability to formulate a mathematical formulation of a problem, based on the formulation in the language of the subject field, to check the correctness of the formulation, including in conditions of uncertainty.
SC 2. The ability to choose, develop and investigate a mathematical analytical or numerical method for solving practical problems, which ensures the required accuracy and reliability of the result.
SC 3. The ability to choose, develop, research and apply mathematical methods to solve practical problems of modeling, design, management, forecasting, decision-making.
SC 4. Ability to develop algorithms for the analysis of uncertain big data, develop appropriate software tools and documentation, design software systems, databases and knowledge bases.
SC 5. The ability to conduct mathematical and computer modeling and computing experiments, collect, visualize, analyze and process the received data, solve formalized problems with the help of specialized software tools.
SC 10. The ability to select, develop, research and apply mathematical models and methods for intellectual analysis of data under conditions of uncertainty.
SC 11. Ability to develop, investigate and apply mathematical methods and algorithms of machine learning, soft computing and computational intelligence for the analysis of uncertain data, forecasting and decision making.
SC 12. Ability to develop and operate specialized software tools for intellectual analysis of data, texts, signals and images.
SC 14. The ability to use modern information technologies for intelligent data analysis, forecasting, decision-making, information search and knowledge extraction.

Learning outcomes

LO 1. Demonstrate knowledge and understanding of basic concepts, principles, theories of fundamental and applied mathematics and use them in practice.
LO 2. Be able to formalize problems formulated in the language of a specific subject area and choose a rational solution method; solve problems by analytical or numerical methods, evaluate the accuracy and reliability of the obtained results and perform their interpretation.
LO 3. To have the methods of development, research and application of mathematical models of complex objects and processes, including with the use of methods of computational intelligence.
LO 5. Build algorithms that are effective in terms of calculation accuracy, stability, speed, and consumption of system and computing resources for numerical research of mathematical models and data analysis, decision-making.
LO 6. To be able to choose, develop and research methods and algorithms for solving mathematical problems of system optimization, operations research, optimal control and decision-making.
LO 7. To be able to apply modern technologies of programming and software development, software implementation of numerical and symbolic algorithms.
LO 12. Know and understand modern methods of solving mathematical problems, statistical and intellectual data analysis, forecasting, etc.
LO 14. To be able to apply existing existing and develop new algorithms and software tools for statistical and intellectual analysis of uncertain data.
LO 15. Be able to apply existing and develop new algorithms and software tools for data, text, signal and image processing.
LO 16. To be able to apply modern information technologies and software for processing large data sets based on distributed and cloud services.

Student workload

The total volume of the course is 180 hours (6 ECTS credits): lectures – 32 hours, laboratory classes – 48 hours, self-study – 100 hours.

Course prerequisites

Bachelos's level of education.

Features of the course, teaching and learning methods, and technologies

Programming skills are required. Study materials are available to students on the teacher's website.

Program of the course

Topics of the lectures

Topic 1. Introduction to the subject.
Topic 2. Construction of features for describing graphs.
Topic 3. Construction of features for describing graphs. Weisfeiler-Lehman Kernel.
Topic 4. Overview of the main elements of convolutional neural networks.
Topic 5. Graph neural networks (GNN).
Topic 6. Mechanism of attention (attention).
Topic 7. Graph neural networks. An overview of GNNs.
Topic 8. General GNN training pipeline. Graph manipulation in GNN.
Topic 9. Analysis of expressiveness of GNN.
Topic 10. Machine learning with heterogeneous graphs.
Topic 11. Knowledge graphs.
Topic 12. Reasoning based on knowledge graphs.
Topic 13. Fast mapping of neural subgraphs.
Topic 14. Deep generative models for graphs.
Topic 15. Transformer architecture in graph neural networks.
Topic 16. Review of the course and discussion of the achieved results.

Topics of the laboratory classes

Topic 1. Getting to know the main libraries for working with graphs in the Python environment.
Topic 2. Calculation of features of graph nodes.
Topic 3. Training embeddings for graph nodes.
Topic 4. Construction and training of GNN.
Topic 5. Construction and training of GraphSAGE and GAT.

Self-study

During self-stury, students study lecture material, prepare for tests, and exams.

Non-formal education

In non-formal education according to the relevant Regulation (http://surl.li/pxssv), the educational component or its individual topics can be taken into account in case of independent completion of professional courses/trainings, obtaining civic education, online education, professional internship, etc.

In particular, individual topics of this component may be taken into account upon successful completion of the following courses:

http://cs224w.stanford.edu/
https://www.udemy.com/course/graph-neural-network/

Course materials and recommended reading

Main literature

1. CS224W: Machine Learning with Graphs.
http://cs224w.stanford.edu/
2. William L. Hamilton. Graph Representation Learning Book.
https://www.cs.mcgill.ca/~wlh/grl_book/
3. DeepWalk: Online Learning of Social Representations.
https://arxiv.org/pdf/1403.6652.pdf
4. Node2vec: Scalable Feature Learning for Networks.
https://arxiv.org/pdf/1607.00653.pdf
5. Network Embedding as Matrix Factorization.
https://arxiv.org/pdf/1710.02971.pdf

Additional literature

6. Semi-Supervised Classification with Graph Convolutional Networks.
https://arxiv.org/pdf/1609.02907.pdf
7. Design Space of Graph Neural Networks.
https://arxiv.org/pdf/2011.08843.pdf
8. Inductive Representation Learning on Large Graphs.
https://arxiv.org/abs/1706.02216
9. Graph Attention Networks.
https://arxiv.org/abs/1710.10903
10. Hierarchical Graph Representation Learning with Differentiable Pooling.
https://arxiv.org/pdf/1806.08804.pdf
11. How Powerful Are Graph Neural Networks?
https://arxiv.org/pdf/1810.00826.pdf
12. Modeling Relational Data with Graph Convolutional Networks.
https://arxiv.org/pdf/1703.06103.pdf
13. Heterogeneous Graph Transformer.
https://arxiv.org/pdf/2003.01332.pdf
14. Complex Embeddings for Simple Link Prediction.
https://arxiv.org/pdf/1606.06357.pdf
15. Embedding Logical Queries on Knowledge Graphs.
https://arxiv.org/pdf/1806.01445.pdf

Assessment and grading

Criteria for assessment of student performance, and the final score structure

A necessary condition for passing the test or exam is the completion of laboratory work.
30 points are awarded for writing control tests.
Passing laboratory tests - 30 points.
Exam – 40 points.

Grading scale

Total points	National	ECTS
90–100	Excellent	A
82–89	Good	B
75–81	Good	C
64–74	Satisfactory	D
60–63	Satisfactory	E
35–59	Unsatisfactory (requires additional learning)	FX
1–34	Unsatisfactory (requires repetition of the course)	F

Norms of academic integrity and course policy

The student must adhere to the Code of Ethics of Academic Relations and Integrity of NTU "KhPI": to demonstrate discipline, good manners, kindness, honesty, and responsibility. Conflict situations should be openly discussed in academic groups with a lecturer, and if it is impossible to resolve the conflict, they should be brought to the attention of the Institute's management.
Regulatory and legal documents related to the implementation of the principles of academic integrity at NTU "KhPI" are available on the website: http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/

Approval

Approved by

Date, signature
31.08.2023

Head of the department
Olena AKHIIEZER

Date, signature
31.08.2023

Guarantor of the educational program
Leonid LYUBCHYK